ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Adversarial Transferred Data-Assisted Soft Sensor for Enhanced Multigrade Quality Prediction

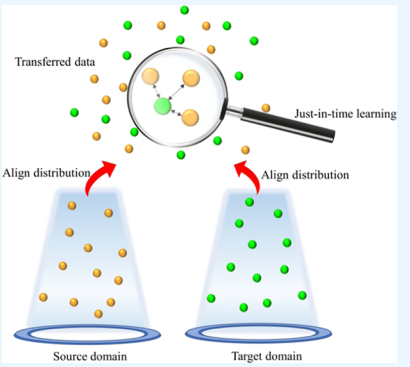
Yun Dai, Chao Yang, Jialiang Zhu, and Yi Liu*

Cite This: *ACS Omega* 2023, 8, 19900−19911

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Although recent transfer learning soft sensors show promising applications in multigrade chemical processes, good prediction performance mainly relies on available target domain data, which is difficult to achieve for a start-up grade. Additionally, only employing a single global model is inadequate to characterize the inner relationship of process variables. A just-in-time adversarial transfer learning (JATL) soft sensing method is developed to enhance multigrade process prediction performance. The distribution discrepancies of process variables between two different operating grades are first reduced by the ATL strategy. Subsequently, by applying the just-in-time learning approach, a similar data set is selected from the transferred source data for reliable model construction. Consequently, with the JATL-based soft sensor, quality prediction of a new target grade is implemented without its own labeled data. Experimental results on two multigrade chemical processes validate that the JATL method can give rise to the improvement of model performance.



1. INTRODUCTION

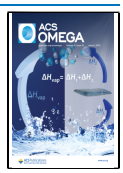
In modern industries, multigrade processes have been widely operated to satisfy the needs of flexible manufacturing and diversified market. By altering the component ratio of raw materials, different product grades that naturally fall under the same chemical mechanism are produced.[1−3] However, the key quality variables of the polyethylene process (melt index, MI) are obtained by offline laboratory analysis instead of online measurement. In such a situation, a large measurement delay occurs, which is not beneficial to process quality control and optimization. Alternatively, soft sensing methods are developed to alleviate the problem.[4−16] The concerned hard-to-measure quality variables are estimated with the help of process variables. As known, adequate training samples are a key factor in reliable model construction. However, for multigrade processes, enough labeled samples often exist in a few specific operating grades, while most of the grades can only collect limited or even no labeled data. Also, a large distribution gap exists between different grades. Therefore, for grades without labeled samples, the construction of an appropriate soft sensor is intractable but necessary.

As a branch of machine learning, transfer learning (TL) aims to tackle the problem of insufficient training data between domains.[17−22] By leveraging knowledge from the related domain having relatively sufficient annotated samples, the predicted target domain having limited labeled samples is expanded. Recently, with good feature extraction capability, deep transfer learning methods[23−26] attempt to transfer knowledge. To address the problem of model construction of multigrade processes with limited samples, deep transfer learning-based soft sensors have been developed. Adversarial transfer learning (ATL)[17] reduces the distribution discrepancy between domains. With the assistance of labeled target data, the transfer learning method enhances the prediction performance.

However, existing TL soft sensors focus on global models, which may be inadequate to learn all of the process characteristics. As illustrated in ref 27, in fact, most soft sensor methods have a finite lifespan. Therefore, flexible architecture with an adaptable structure tends to be appealing and appropriate for chemical processes. To this end, researchers applied the just-in-time learning (JITL) approaches to several local modeling tasks.[2,28−30] As query comes, some relevant samples are picked for the establishment of a local prediction model, which is suitable to track the process characteristics. To our best knowledge, the combination of JITL and TL to soft sensing of multigrade processes has rarely been reported. Moreover, based on ATL, JITL strategy can reduce the dependence on labeled samples in the target domain.

This work develops a just-in-time adversarial transfer learning (JATL) soft sensing model to predict multigrade processes online. First, to reduce the distribution discrepancies

Received: March 17, 2023
Accepted: May 12, 2023
Published: May 25, 2023



https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

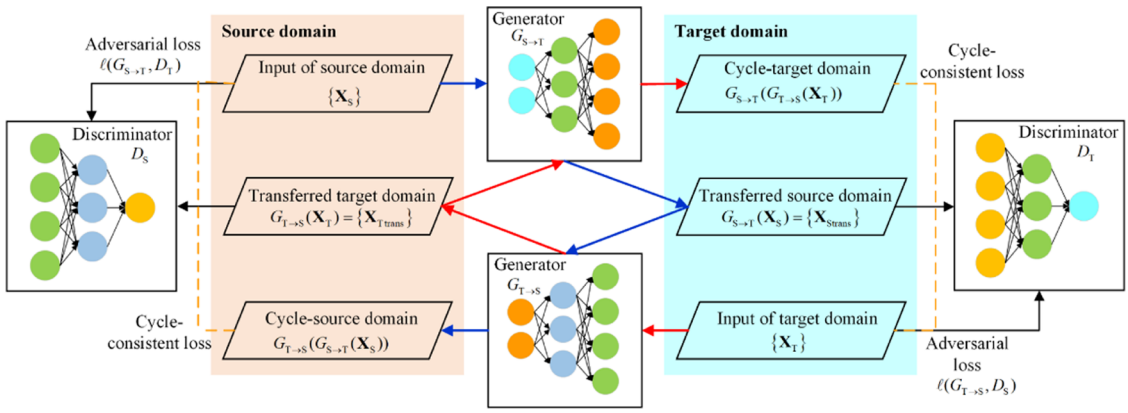


Figure 1. Structure of ATL strategy.

across grades, the ATL strategy is adopted among process variables. Virtual target labeled data are augmented by the transferred labeled source domain data. Subsequently, the JITL strategy is applied to select similar transferred source data according to a query target sample for model construction. Without loss of generality, two common methods, namely, regularized extreme learning machine (RELM)[31] and least-squares support vector regression (LSSVR),[32] are adopted for the construction of JATL-based soft sensing models. The JATL framework can also be combined with other soft sensors for multigrade processes. By utilizing the source domain information, it is expected that the prediction accuracy of the target domain is enhanced.

The remaining sections are structured as follows. Some preliminaries are concisely reviewed in Section 2. In Section 3, the implementations of JATL are described. In Section 4, two multigrade processes are evaluated to validate effectiveness. Finally, Section 5 makes a conclusion.

2. PRELIMINARIES

2.1. Generative Adversarial Network. Recently, as a popular deep generative model, generative adversarial network (GAN)[33] can generate virtual samples similar to real data. In general, discriminator D and generator G form the GAN network. A mini-max game is played between the G network and the D network. Specifically, the trained G network attempts to produce samples similar to the real data and deceive the D network. Meanwhile, the trained D network makes an effort to improve the ability to discriminate between virtual and real data. The GAN's adversarial training procedure is described as follows:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (1)$$

where $\mathbb{E}$ denotes the expectation. $V(D, G)$ represents the objective function. $\mathbf{x}$ represents the real samples with distribution of $p_{\text{data}}(\mathbf{x})$. The input to the G network is the Gaussian noise $\mathbf{z}$ with distribution of $p_z(\mathbf{z})$. $G(\mathbf{z})$ denotes the generated virtual data.

2.2. RELM Modeling Method. As a typical neural network with a simple structure, RELM[31] shows its fast and good modeling properties. RELM aims to learn a function mapping f: $\mathbf{X} \rightarrow \mathbf{Y}$ depending on the labeled modeling data set $\{\mathbf{S}\} = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1,\cdots,r}$. The mathematical model of RELM is defined as

$$y_i = \sum_{j=1}^{L} \beta_j h(\omega_j^{\mathrm{T}}\mathbf{x}_i + \mathbf{b}_j)\ i = 1, \cdots, r \quad (2)$$

where L, $\boldsymbol{\beta}$, ω, $\mathbf{b}$, and r, and $h(\cdot)$ denote the hidden-node numbers, the output weight of the hidden layer, the input weight of the hidden node, the bias, the labeled-sample numbers, and the activation function, respectively.

The closed-form solution of $\boldsymbol{\beta}$ is expressed as

$$\hat{\beta} = \mathbf{H}^{+}\mathbf{Y} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{Y} \quad (3)$$

To efficiently determine the RELM method's parameters, namely, the regularization parameter λ and the hidden-node number, the fast leave-one-out (FLOO) criterion is utilized.[18] Hence, the predicted output $\hat{y}_t$ of a test sample $\mathbf{x}_t$ is obtained by the RELM model as

$$\hat{y}_t = h(\mathbf{x}_t)\hat{\beta} = h(\mathbf{x}_t)(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{Y} \quad (4)$$

2.3. LSSVR Modeling Method. LSSVR[32] is a kernel-based modeling method defined to minimize

$$\begin{cases} \min J(\mathbf{w}, b, \mathbf{e}) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\|\mathbf{e}\|^2 \\ \text{s. t.} \quad y_k - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_k) - b - e_k = 0,\ k = 1, \cdots, r \end{cases} \quad (5)$$

where $\mathbf{e} = [e_1, e_2, \cdots, e_r]^{\mathrm{T}}$ denotes the modeling error. $\mathbf{w}$ denotes the model weight parameter vector. b denotes the bias term. $\gamma(\gamma \geq 0)$ denotes the regularization parameter. ϕ denotes a feature map.

Here, the FLOO strategy is also applied to the parameter selection of LSSVR and the model estimation output $\hat{y}_t$ of a sample $\mathbf{x}_t$ can be obtained

$$\hat{y}_t = f(\mathbf{w}, b, \mathbf{x}_t) = \sum_{i=1}^{r} \alpha_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_t) \rangle + b \quad (6)$$

where $[\alpha_1, \cdots, \alpha_r]^{\mathrm{T}}$ are the Lagrange multipliers which can be obtained by solving eq 5.[32]

3. JUST-IN-TIME TRANSFER LEARNING MODELING APPROACH

3.1. Problem Statement. Generally, traditional soft sensor models such as RELM and LSSVR are constructed for a labeled set $\{\mathbf{S}\} = \{\mathbf{X}, \mathbf{Y}\}$. For multigrade processes, in

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

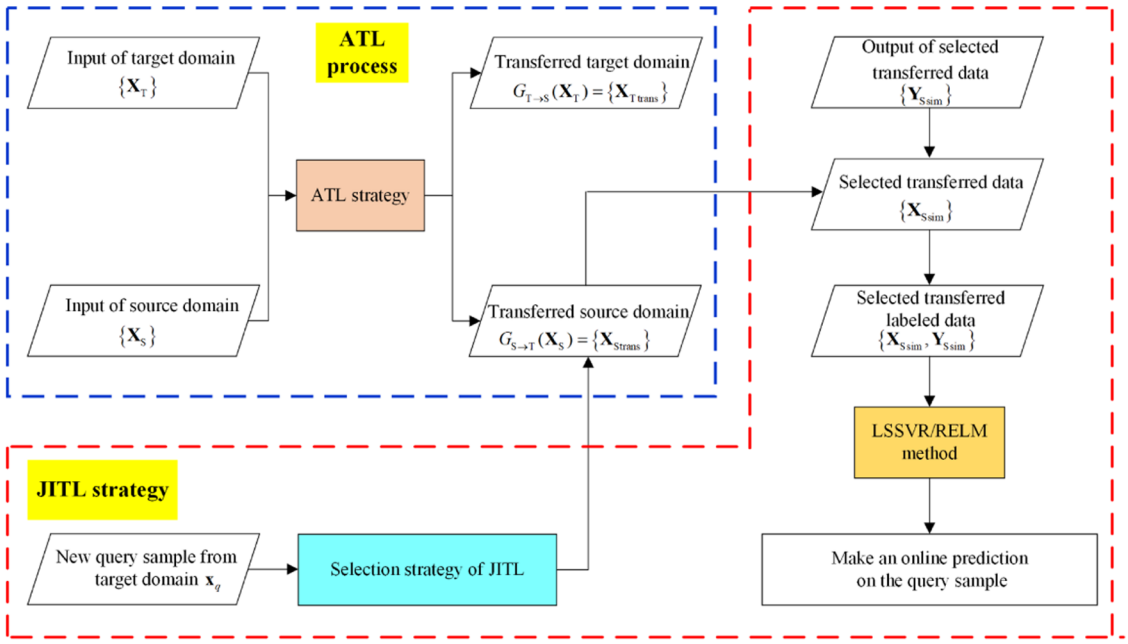


Figure 2. Flowchart of JATL-based soft sensor.

some cases, no labeled samples are obtained for some specific operating grades in their initial operating periods, which hinders the construction of an accurate prediction model. To this end, a JATL-based online soft sensing framework is proposed for quality inferring in multigrade processes. Here, the source domain having relatively sufficient labeled data is represented as $\{\mathbf{X}_S, \mathbf{Y}_S\}$, where $\{\mathbf{X}_S\} = \{\mathbf{X}_{Si}\}_{i=1}^{M}$ denotes the input and $\{\mathbf{Y}_S\} = \{y_{Si}\}_{i=1}^{M}$ denotes the output with M samples. The target domain without labeled quality measurements is denoted as $\{\mathbf{X}_T\} = \{\mathbf{x}_{Tj}\}_{j=1}^{N}$ with N samples. In such a situation, for the target domain, the construction of an accurate soft sensor is intractable. Using the TL strategy in soft sensors, useful information from the source domain is utilized. Furthermore, the JITL strategy makes soft sensors flexible. Consequently, the advantages of TL and JITL can be exploited simultaneously by the JATL strategy.

3.2. Transfer Learning Model Based on Adversarial Domain Adaptation. As a framework for unsupervised learning, ATL attempts to align the distributions between domains. As shown in Figure 1, a forward GAN and a backward one make up the ATL. The loss functions for two GANs are formulated as follows:[17]

$$l(G_{S\to T}, D_T) = \mathbb{E}_{\mathbf{x}_T\sim p(\mathbf{x}_T)}[(D_T(\mathbf{x}_T) - 1)^2] + \mathbb{E}_{\mathbf{x}_S\sim p(\mathbf{x}_S)}[D_T(G_{S\to T}(\mathbf{x}_S))^2] \quad (7)$$

$$l(G_{T\to S}, D_S) = \mathbb{E}_{\mathbf{x}_S\sim p(\mathbf{x}_S)}[(D_S(\mathbf{x}_S) - 1)^2] + \mathbb{E}_{\mathbf{x}_T\sim p(\mathbf{x}_T)}[D_S(G_{T\to S}(\mathbf{x}_T))^2] \quad (8)$$

where $G_{S\to T}$ and $G_{T\to S}$ represent the forward and backward generator networks, respectively. D_T and D_S represent the discriminator networks in the target and source domains, respectively. Here, the least-square loss function replaces the commonly used cross-entropy function, due to its vanishing gradients problem.

Additionally, to guarantee the preservation of local variable information in the process of distribution discrepancy reduction, a cycle-consistent loss $l(G_{S\to T}, G_{T\to S})$ is employed. Similarly, the least-squares loss function is utilized in $l(G_{S\to T}, G_{T\to S})$, expressed as

$$l(G_{S\to T}, G_{T\to S}) = \mathbb{E}_{\mathbf{x}_S\sim p(\mathbf{x}_S)}[(G_{T\to S}(G_{S\to T}(\mathbf{x}_S)) - \mathbf{x}_S)^2] + \mathbb{E}_{\mathbf{x}_T\sim p(\mathbf{x}_T)}[(G_{S\to T}(G_{T\to S}(\mathbf{x}_T)) - \mathbf{x}_T)^2] \quad (9)$$

Consequently, the total optimization function is obtained as follows:

$$l_{total} = l(G_{S\to T}, D_T) + l(G_{T\to S}, D_S) + l(G_{S\to T}, G_{T\to S}) \quad (10)$$

When the training process of ATL reaches convergence, the process variables of the source domain are transferred to the target one and are denoted as $\{\mathbf{X}_{S_{trans}}\}$, while the quality variables remain unchanged as $\{\mathbf{Y}_S\}$. Meanwhile, the process variables in the target domain can be transferred to the source domain, denoted as $\{\mathbf{X}_{T_{trans}}\}$. Here, we choose the transferred source domain for model construction instead of the transferred target domain. The reason for choosing the transferred source domain is that the useful information in the source domain is expected to be transferred to the concerned target domain using transfer learning. Moreover, the transferred source domain is more convenient for online prediction than the transferred target domain. This is mainly because the former only needs to perform one transfer step from the source domain to the target domain, while the latter needs to perform multiple transfer learning steps for target test samples. The marginal probability distributions of $\{\mathbf{X}_{S_{trans}}\}$ and $\{\mathbf{X}_T\}$ are similar, and the distribution discrepancy is reduced.

3.3. JATL-Based Soft Sensor Framework. It is challenging to directly apply a global model for complex multigrade processes. Furthermore, the obtained training samples are insufficient to characterize all of the process properties. To alleviate the aforementioned problems, the JITL methods have been developed for online modeling. When a

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

Table 1. Comparisons of Four RELM-Based Prediction Methods for CSTR[a]

	RMSE/MAE			
	RELM	JRELM	ATL-RELM	JATL-RELM
G1 → G2	0.730/0.570	**0.677/0.511**	0.954/0.816	0.894/0.782
G1 → G3	12.073/6.454	4.088/3.049	0.905/1.319	**0.724/1.126**
G2 → G1	0.902/0.740	**0.418/0.326**	1.689/0.954	1.302/1.028
G2 → G3	19.984/15.339	1.446/1.285	1.297/1.367	**0.663/1.112**
G3 → G1	100.81/71.965	1.523/1.296	0.636/0.494	**0.515/0.415**
G3 → G2	83.467/62.030	2.009/1.555	1.129/0.895	**0.757/0.637**

[a]Values in bold indicate better prediction results.

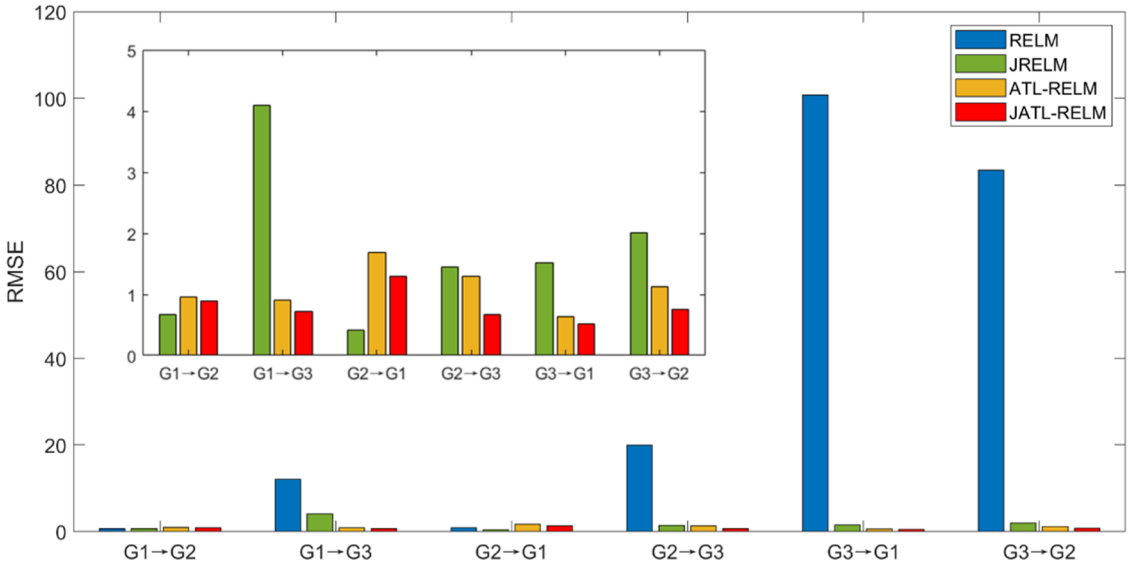


Figure 3. RMSE value comparisons of four RELM-based methods for CSTR.

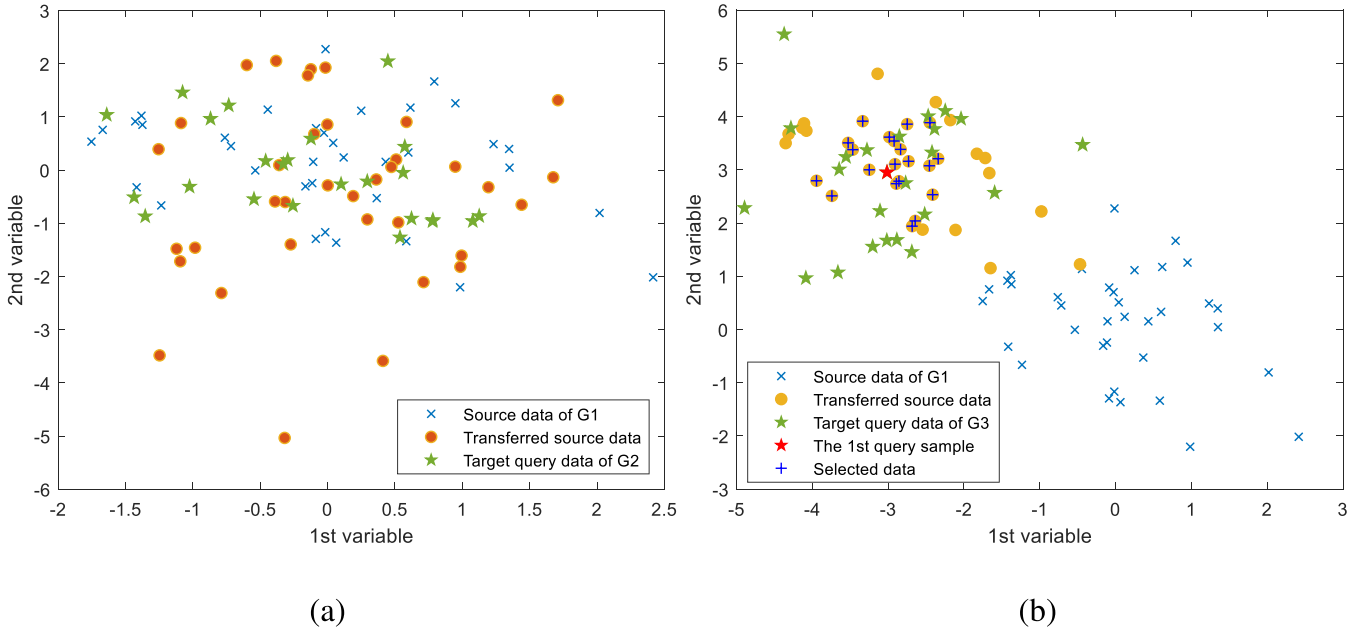


Figure 4. Distribution scatter comparison between different domains for CSTR: (a) G1 and G2, (b) G1 and G3.

query sample $\mathbf{x}_q$ from the target domain comes, three steps are involved in the JITL-based mode. First, select the most similar samples from $\{\mathbf{X}_{S_{trans}}, \mathbf{Y}_S\}$ as a new set. Second, construct an online soft sensor model such as RELM or LSSVR with the selected samples. Third, give a prediction on the output $\hat{y}_q$. Repeating the aforementioned three-step procedures, another RELM or LSSVR prediction model is established for a new query sample.

A general measurement rule, namely, Euclidean distance, is applied in the JITL process to measure the similarity between

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

Figure 5. Performance of four RELM-based methods for CSTR.

samples. The similarity factor sim_{qi} between $\mathbf{x}_i$ and $\mathbf{x}_q$ is formulated as[29,30]

$$\mathrm{sim}_{qi} = \exp(-\left\|\mathbf{x}_q - \mathbf{x}_i\right\|),\ i = 1, \cdots, r \tag{11}$$

where the value sim_{qi} varies between 0 and 1. The closer sim_{qi} is to 1, the more similar $\mathbf{x}_q$ is to $\mathbf{x}_i$. According to eq 11, a similar data set $\{\mathbf{X}_{S_{\mathrm{sim}}}, \mathbf{Y}_{S_{\mathrm{sim}}}\}$ with n_{sim} samples is collected, where n_{sim} is determined by the cumulative similarity factor[2] defined as below:

$$S_q = \frac{\sum_{i=1}^{n_{\mathrm{sim}}} \mathrm{sim}_{qi}}{\sum_{i=1}^{M} \mathrm{sim}_{qi}},\ n_{\mathrm{sim}} \le M \tag{12}$$

This index measures the cumulative similarity of the n_{sim} most similar samples to the total transferred source domain data. For example, the choice of $S_q = 0.9$ means 90% of the most similar samples are selected. Then, it can determine the number of the most similar samples in practical applications. Although other similarity criteria[28] can also be applied to build a JATL-based soft sensor, it is not our main purpose.

Combing the aforementioned strategy of ATL and JITL into a unified framework, a JATL-based online soft sensor is developed for the multigrade processes. It constructs the real-time model while utilizing the information from the other domains. The modeling scheme of the JATL-based method is shown in Figure 2. First, the ATL strategy is used to align

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

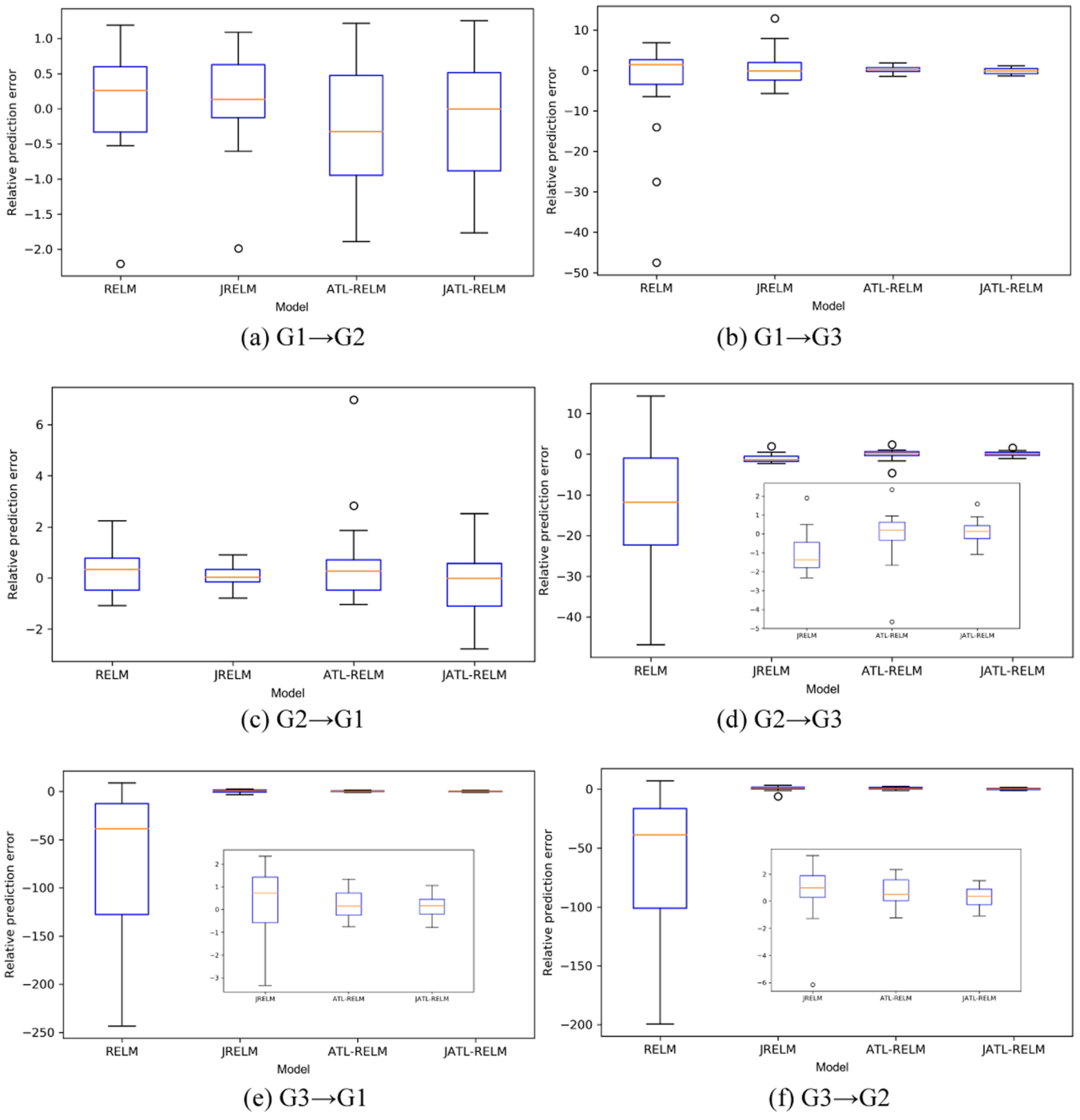


Figure 6. Relative prediction errors of four RELM-based methods for CSTR.

Table 2. Comparisons of Four LSSVR-Based Prediction Methods for CSTR[a]

	RMSE/MAE			
	LSSVR	JLSSVR	ATL-LSSVR	JATL-LSSVR
G1 → G2	0.621/0.493	**0.611/0.477**	0.915/0.795	0.912/0.788
G1 → G3	2.871/2.807	1.911/1.766	0.752/0.585	**0.692/0.534**
G2 → G1	0.486/0.401	**0.471/0.395**	1.392/1.144	1.376/1.138
G2 → G3	2.043/1.930	1.290/1.033	1.253/0.798	**0.926/0.776**
G3 → G1	1.408/1.305	0.950/0.774	0.579/0.486	**0.521/0.432**
G3 → G2	1.696/1.571	1.563/1.358	1.066/0.866	**0.906/0.764**

[a]Values in bold indicate better prediction results.

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

Table 3. Comparisons of Four RELM-Based Prediction Methods for Polyethylene[a]

	RMSE/MAE			
	RELM	JRELM	ATL-RELM	JATL-RELM
G1 → G2	2.638/2.114	2.048/1.711	1.935/1.570	**1.868/1.510**
G1 → G3	79.389/74.693	40.465/32.694	25.085/18.950	**22.607/18.127**
G2 → G1	12.820/10.943	10.089/8.917	10.8860/9.932	**9.919/8.691**
G2 → G3	31.321/24.581	25.990/21.262	24.540/19.825	**23.456/19.242**
G3 → G1	34.189/29.098	27.382/22.396	13.251/10.386	**8.929/7.040**
G3 → G2	29.505/24.443	6.316/3.884	1.845/**1.474**	**1.799**/1.493

[a]Values in bold indicate better prediction results.

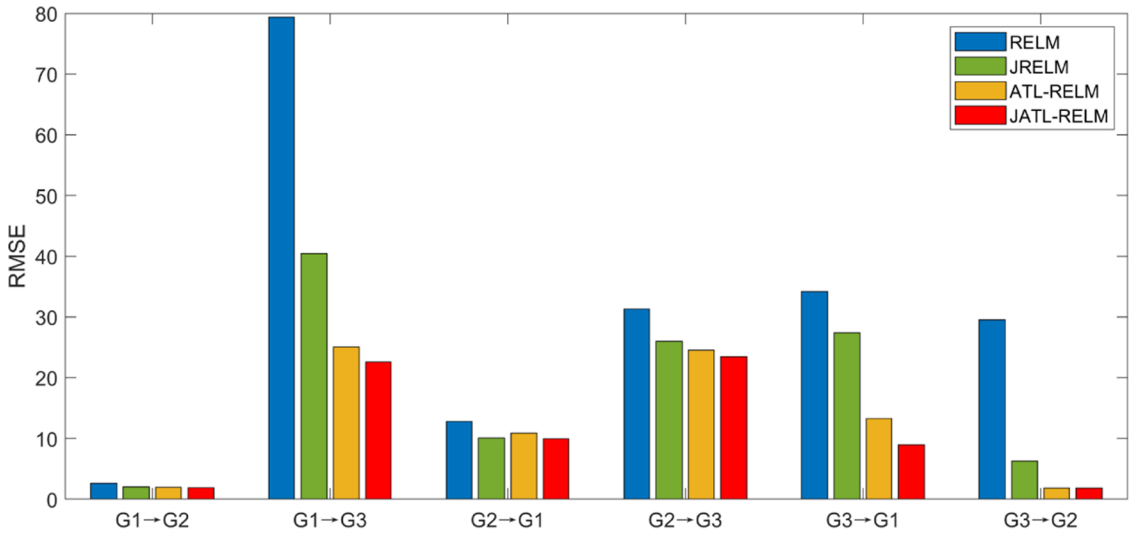


Figure 7. RMSE value comparisons of four RELM-based methods for polyethylene.

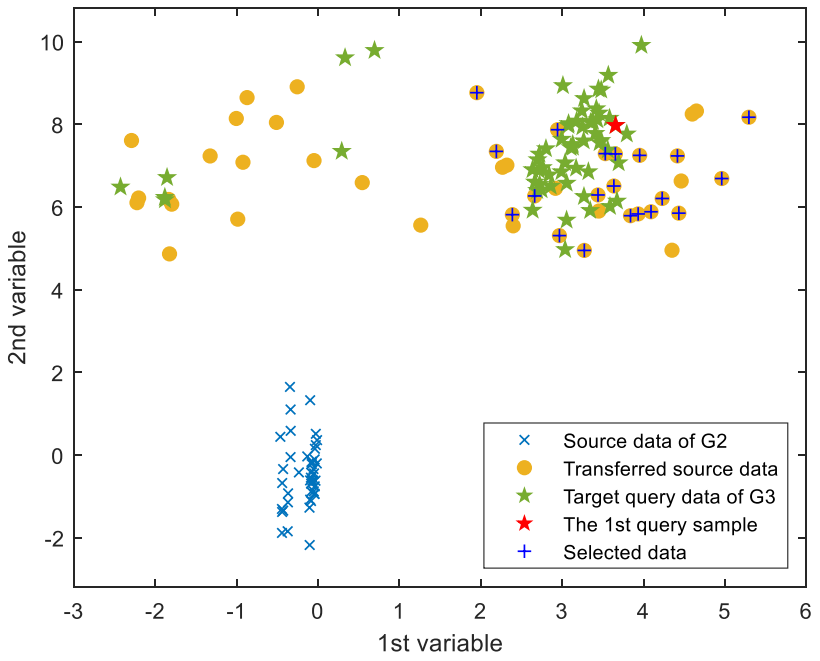


Figure 8. Distribution scatter comparison between G2 and G3 for polyethylene.

domain distribution. Then, the source process samples are transferred to the target domain denoted as $\{\mathbf{X}_{S_{trans}}, \mathbf{Y}_S\}$, which follows the same distribution with the target domain. Subsequently, according to the Euclidean distance-based selection criterion, a similar data set $\{\mathbf{X}_{S_{sim}}, \mathbf{Y}_{S_{sim}}\}$ from the transferred source data $\{\mathbf{X}_{S_{trans}}, \mathbf{Y}_S\}$ is selected for a query sample $\mathbf{x}_q$. Based on $\{\mathbf{X}_{S_{sim}}, \mathbf{Y}_{S_{sim}}\}$, an online RELM or LSSVR model is constructed to perform prediction on the target query sample. Consequently, compared to the fixed global transfer learning models, JATL-based methods are more flexible.

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

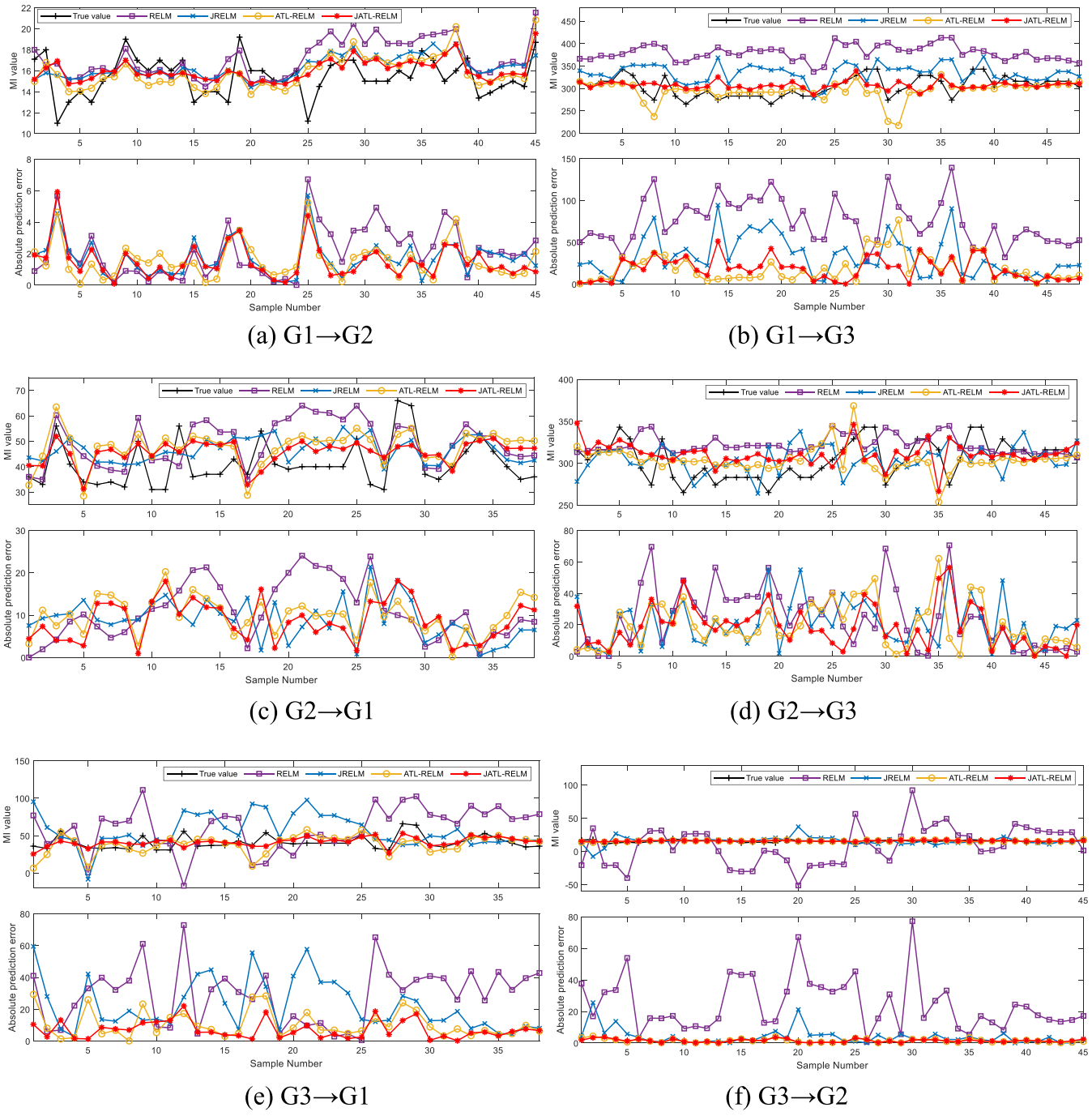


Figure 9. Performance of four RELM-based methods for polyethylene.

4. RESULTS AND DISCUSSION

To verify the feasibility of the JATL strategy, it is conducted on two multigrade chemical cases. RELM and LSSVR are adopted to form JATL-RELM and JATL-LSSVR models. Meanwhile, other models, i.e., RELM/LSSVR, JRELM/JLSSVR, and ATL-RELM/ATL-LSSVR, are constructed for performance comparison. Root-mean-square error (RMSE) and mean absolute error (MAE) indices are used

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{t=1}^{m}|y_t - \hat{y}_t| \tag{14}$$

where $\hat{y}_t$ and y_t are the predicted and real values with m test samples.

4.1. Simulated Process. Continuous stirred tank reactor (CSTR) is widely employed in the polymerization process to produce multigrade products. The mechanism of the CSTR process can be found in ref 34. To give a prediction on the quality variable reactor concentration, two process variables are selected as inputs.[2,17,18] For convenience, three steady-state grades with 60 labeled samples are explored. Among them, 36 samples are for training, and 24 samples are for testing. Each

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

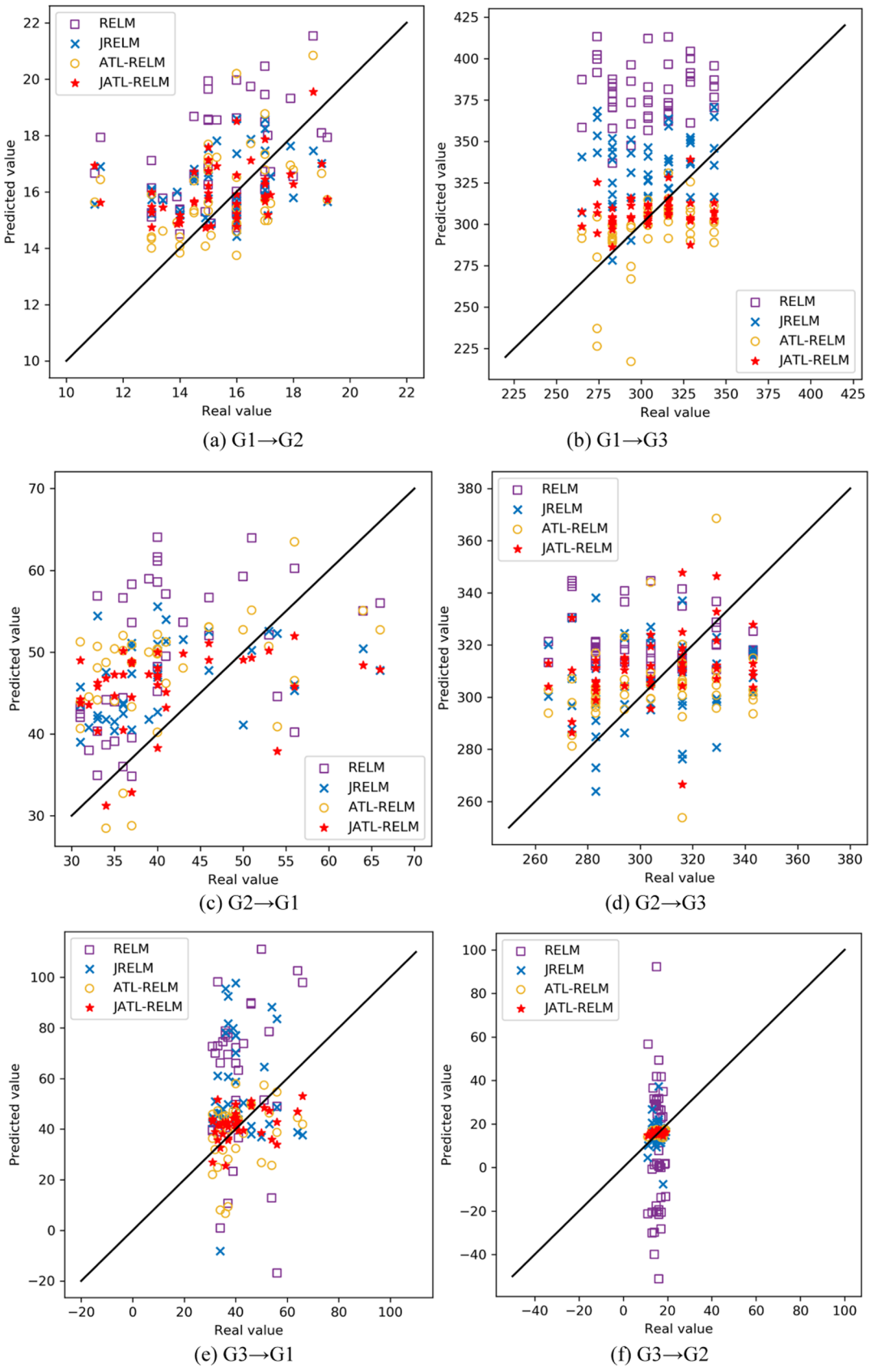


Figure 10. Prediction scatters of four RELM-based methods for polyethylene.

scenario assumes that only one grade has sufficient labeled samples, while other grades only have unlabeled training samples.

The detailed prediction results of RELM, JRELM, ATL-RELM, and JATL-RELM methods are listed in Table 1. Intuitively, RMSE value comparisons are shown in Figure 3.

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

Table 4. Comparisons of Four LSSVR-Based Prediction Methods for Polyethylene[a]

	RMSE/MAE			
	LSSVR	JLSSVR	ATL-LSSVR	JATL-LSSVR
G1 → G2	3.213/2.568	2.986/2.415	2.269/1.828	**2.078/1.613**
G1 → G3	89.443/76.058	44.213/38.778	31.471/24.430	**28.110/23.132**
G2 → G1	16.859/12.234	11.821/10.394	12.228/11.250	**10.945/10.002**
G2 → G3	34.498/28.994	27.322/22.316	27.300/21.239	**21.504/16.878**
G3 → G1	40.027/16.878	36.847/25.327	16.631/12.222	**13.374/10.321**
G3 → G2	6.922/6.490	4.006/2.478	2.979/2.260	**2.064/2.260**

[a]Values in bold indicate better prediction results.

For the 1st scenario, grade 1 (i.e., G1) is taken as S, while grade 2 (i.e., G2) and grade 3 (i.e., G3) are taken as T. For G1 → G2, RELM and JRELM achieve higher prediction accuracy than ATL-RELM and JATL-RELM methods. This is mainly caused by the small distribution discrepancy between G1 and G2, as described in refs 17 and 18. The source data of G1, transferred source data by ATL strategy, and target query data of G2 are plotted in Figure 4a. The small distribution gap between the source data and the target query data weakens the improvement of the model performance by ATL strategy. It should be noted that benefit from the JITL strategy, JRELM is superior to RELM, and JATL-RELM is more accurate than ATL-RELM. For scenario G1 → G3, the JATL-RELM has the smallest RMSE and MAE values, which indicates that JATL-RELM outperforms the other three methods. Figure 4b shows the scatter plot of the source data of G1, transferred data, and query data of G3. For the first query sample, 20 relevant samples are chosen from the transferred data of G1 by JITL strategy. Different from the distribution of all of the transferred data, selected samples are closer to the query sample. In the second scenario, G2 is taken as S. Similar conclusions are drawn from the prediction results. In the last scenario of G3 taken as S, the JATL-RELM method achieves the best accuracy by selecting the most similar data from the transferred G3.

Due to the inferior prediction accuracy of RELM method, Figure 5 only shows the comparison results of JRELM, ATL-RELM, and JATL-RELM methods. In the two cases between G1 and G2, the prediction results of JATL-RELM are inferior to the JRELM method due to the large similarity between the two grades. However, in other cases, the predicted output of JATL-RELM tracks the real output values well. It illustrates that the online JATL-RELM method shows more accurate prediction results than others. The relative prediction errors ($\delta = (\hat{y}_t - y_t)/y_t$) are illustrated using box plots in Figure 6. For G1 and G3, G2, and G3, the quartile and median values in the box plots reveal that the JATL-RELM method achieves the best prediction performance.

Additionally, to further illustrate the JATL-based soft sensor framework for the CSTR process, the JATL-LSSVR model is built. In different scenarios, RMSE and MAE values of four soft sensing methods are tabulated in Table 2. Except for the scenarios between G1 and G2, the prediction performance of JATL-LSSVR is the most superior. The main reason lies in the source, and the concerned target domains are distributed similarly using the ATL strategy. Furthermore, the most relevant samples to the query one assist in building a more accurate LSSVR model.

4.2. Industrial Polyethylene Process. The JATL strategy is further investigated using an industrial polyethylene process. In total, 266 labeled samples with their related input variables are obtained in three grades, with 83, 90, and 93 ones in G1, G2, and G3, respectively.[2,17,18] From each grade, 45 labeled samples are regarded as training sets. In any online transfer learning scenario, only the source grade has relatively enough labeled training samples, while the target grade only contains unlabeled data. The process characteristics of the three grades are different, and product quality (i.e., MI) values vary widely, i.e., G1 (30−50), G2 (10−20), and G3 (240−350).

In the three transfer learning scenarios, the RMSE values and MAE values of four methods (i.e., JATL-RELM, RELM, JRELM, and ATL-RELM) are listed in Table 3. For visualization, the histograms of the RMSE values are shown in Figure 7. After applying the ATL strategy, the ATL-RELM method shows better prediction accuracy than RELM and JRELM. Additionally, using the JITL strategy, similar samples are selected from the transferred source domain. Therefore, JATL-RELM outperforms the other three methods with the smallest RMSE and MAE values. Taking G2 → G3 as an example, the distribution scatters of source data and target query data are shown in Figure 8. The distributions of G2 and G3 differ significantly, which makes it impossible to predict the query samples in G3 with the model trained by the labeled samples in G2. By the ATL strategy, the distribution gap between G2 and G3 is reduced. For the first query sample, 18 similar samples are chosen from the transferred data by the JITL strategy. The prediction results and errors are shown in Figure 9. Compared with the other methods, the JATL-RELM tracks the trend of MI well. In addition, using the scattergrams shown in Figure 10, the points of the JATL-RELM method lying tighter than the other three methods indicate its better prediction performance.

Moreover, the JATL-LSSVR model is constructed to further show the feasibility of the JATL-based method. As shown in Table 4, JATL-RELM achieves the smallest RMSE and MAE values. This indicates that the combination of ATL and JITL strategies can improve prediction accuracy compared to the method with only one strategy. The reason is that the ATL strategy aligns the distribution between grades, and the JITL strategy selects the most similar samples to establish the local model. According to the comparisons of the two cases, the JATL-based framework achieves the best prediction results for multigrade processes.

5. CONCLUSIONS

A JATL-based soft sensing modeling framework is developed for reliable online prediction of multigrade processes without annotated samples of the target domain. Specifically, the ATL strategy reduces distribution discrepancies between different grades by appropriately transferring process information. Based on the transferred source domain data, a suitable similar data set is collected for the query target sample, and the distribution

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911

gap is further minimized. By constructing the real-time adversarial JATL modeling method, the prediction performance for multigrade processes is enhanced compared to several common methods. Its feasibility and superiority are demonstrated using two multigrade processes.

AUTHOR INFORMATION

Corresponding Author

Yi Liu − *Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310023, People's Republic of China;* orcid.org/0000-0002-4066-689X; Phone: +86-571-8529-0402; Email: yliuzju@zjut.edu.cn

Authors

Yun Dai − *Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310023, People's Republic of China*

Chao Yang − *State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, People's Republic of China*

Jialiang Zhu − *Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310023, People's Republic of China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c01832

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China (Grant Nos. 62022073 and 61873241).

NOMENCLATURE

ATL	adversarial transfer learning
CSTR	continuous stirred tank reactor
FLOO	fast leave-one-out
JATL	just-in-time adversarial transfer learning
JITL	just-in-time learning
LSSVR	least-squares support vector regression
MAE	mean absolute error
MI	melt index
RELM	regularized extreme learning machine
RMSE	root-mean-square-error
TL	transfer learning

SYMBOLS

$\mathbf{b}$	the bias vector
b	model bias term
D_S	discriminator network in source domain of ATL
D_T	discriminator network in target domain of ATL
$\mathbb{E}$	the expectation
$\mathbf{e}$	modeling error
$G(\mathbf{z})$	the virtual data generated by the G network
$G_{S\to T}$	the forward generator network of ATL
$G_{T\to S}$	the backward generator network of ATL
$h(\cdot)$	activation function
L	hidden nodes of RELM model
M	size of source labeled samples
N	size of target labeled samples
n_{sim}	number of similar samples
$p_{data}(\mathbf{x})$	distribution of $\mathbf{x}$
$p_z(\mathbf{z})$	distribution of $\mathbf{z}$
r	number of labeled samples
sim_{qi}	similarity factor between $\mathbf{x}_i$ and sample $\mathbf{x}_q$
S_q	the cumulative similarity factor
$V(D, G)$	objective function
$\mathbf{w}$	the model weight parameter vector of LSSVR
$\mathbf{x}$	real data
$\mathbf{x}_q$	a query sample
$\mathbf{x}_t$	test sample
$\hat{y}_q$	predicted output of $\mathbf{x}_q$
y_t	real value of the test sample
$\hat{y}_t$	predicted output of the test sample
$\boldsymbol{z}$	the Gaussian noise
$\boldsymbol{\beta}$	the output weight vector
γ	regularization parameter
λ	the penalty coefficient
ω	the input weight of hidden node
ϕ	feature map function

REFERENCES

(1) Abeykoon, C. Design and applications of soft sensors in polymer processing: A review. *IEEE Sens. J.* **2019**, *19*, 2801−2813.

(2) Liu, Y.; Chen, J. H. 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.

(3) Luo, Y.; Ierapetritou, M. Multifeedstock and multiproduct process design using neural network surrogate flexibility constraints. *Ind. Eng. Chem. Res.* **2023**, *62*, 2067−2079.

(4) Kadlec, P.; Grbic, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2011**, *35*, 1−24.

(5) Xibilia, M. G.; Patanè, L. Echo-state networks for soft sensor design in an SRU process. *Inf. Sci.* **2021**, *566*, 195−214.

(6) Zhang, B. L.; Han, Y. M.; Li, C. F.; Geng, Z. Q. Novel gray orthogonal echo state network integrating the process mechanism for dynamic soft sensor development. *Ind. Eng. Chem. Res.* **2021**, *60*, 14955−14967.

(7) Alakent, B. Soft-sensor design via task transferred just-in-time-learning coupled transductive moving window learner. *J. Process Control* **2021**, *101*, 52−67.

(8) Guo, R.; Liu, H. A hybrid mechanism- and data-driven soft sensor based on the generative adversarial network and gated recurrent unit. *IEEE Sens. J.* **2021**, *21*, 25901−25911.

(9) Shao, W. M.; Ge, Z. Q.; Yao, L.; Song, Z. H. Bayesian nonlinear Gaussian mixture regression and its application to virtual sensing for multimode industrial processes. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 871−885.

(10) Liu, Y. Q.; Xie, M. Rebooting data-driven soft-sensors in process industries: A review of kernel method. *J. Process Control* **2020**, *89*, 58−73.

(11) Shang, C.; Huang, X. L.; You, F. Q. Data-driven robust optimization based on kernel learning. *Comput. Chem. Eng.* **2017**, *106*, 464−479.

(12) Gao, S.; Dai, Y.; Li, Y. J.; Jiang, Y. X.; Liu, Y. Augmented flame image soft sensor for combustion oxygen content prediction. *Meas. Sci. Technol.* **2023**, *34*, No. 015401.

(13) Jiang, B.; Liu, Y.; Geng, H.; Wang, Y.; Zeng, H.; Ding, J. A holistic feature selection method for enhanced short-term load forecasting of power system. *IEEE Trans. Instrum. Meas.* **2022**, *72*, No. 2500911.

(14) He, Y. L.; Tian, Y.; Xu, Y.; Zhu, Q. X. Novel soft sensor development using echo state network integrated with singular value decomposition: Application to complex chemical processes. *Chemom. Intell. Lab. Syst.* **2020**, *200*, No. 103981.

(15) Lyu, Y. T.; Chen, J. H.; Song, Z. H. Synthesizing labeled data to enhance soft sensor performance in data-scarce regions. *Control Eng. Pract.* **2021**, *115*, No. 104903.
(16) Yang, Z. Y.; Jia, R. N.; Wang, P. L.; Yao, L.; Shen, B. B. Supervised attention-based bidirectional long short-term memory network for nonlinear dynamic soft sensor application. *ACS Omega* **2023**, *8*, 4196−4208.
(17) Liu, Y.; Yang, C.; Zhang, M. T.; Dai, Y.; Yao, Y. Development of adversarial transfer learning soft sensor for multigrade processes. *Ind. Eng. Chem. Res.* **2020**, *59*, 16330−16345.
(18) Liu, Y.; Yang, C.; Liu, K. X.; Chen, B. C.; Yao, Y. Domain adaptation transfer learning soft sensor for product quality prediction. *Chemom. Intell. Lab. Syst.* **2019**, *192*, No. 103813.
(19) Wang, Y. L.; Wu, D. Z.; Yuan, X. F. LDA-based deep transfer learning for fault diagnosis in industrial chemical processes. *Comput. Chem. Eng.* **2020**, *140*, No. 106964.
(20) Li, W. J.; Gu, S.; Zhang, X. P.; Chen, T. Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes. *Comput. Chem. Eng.* **2020**, *139*, No. 106904.
(21) Wang, K.; Zhou, W. X.; Mo, Y. F.; Yuan, X. F.; Wang, Y. L.; Yang, C. H. New mode cold start monitoring in industrial processes: A solution of spatial−temporal feature transfer. *Knowl.-Based Syst.* **2022**, *248*, No. 108851.
(22) Liu, J. X.; Hou, G. Y.; Shao, W. M.; Chen, J. H. A supervised functional Bayesian inference model with transfer-learning for performance enhancement of monitoring target batches with limited data. *Process Saf. Environ. Prot.* **2023**, *170*, 670−684.
(23) Tan, C. Q.; Sun, F. C.; Kong, T.; Zhang, W. C.; Yang, C.; Liu, C. F.In *A Survey on Deep Transfer Learning*, Conference of Artificial Neural Networks and Machine Learning, 2018; pp 270−279.
(24) Chai, Z.; Zhao, C. H.; Huang, B.; Chen, H. T. A deep probabilistic transfer learning framework for soft sensor modeling with missing data. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 7598−7609.
(25) Zhao, X.; Luo, T. F.; Jin, H. Predicting diffusion coefficients of binary and ternary supercritical water mixtures via machine and transfer learning with deep neural network. *Ind. Eng. Chem. Res.* **2022**, *61*, 8542−8550.
(26) Li, W. H.; Huang, R. Y.; Li, J. P.; Liao, Y. X.; Chen, Z. Y.; He, G. L.; Yan, R. Q.; Gryllias, K. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mech. Syst. Signal Process.* **2022**, *167*, No. 108487.
(27) Lu, B.; Chiang, L. Semi-supervised online soft sensor maintenance experiences in the chemical industry. *J. Process Control* **2018**, *67*, 23−34.
(28) Fujiwara, K.; Kano, M.; Hasebe, S.; Takinami, A. Soft-sensor development using correlation-based just-in-time modeling. *AIChE J.* **2009**, *55*, 1754−1765.
(29) Jin, H. P.; Li, J. G.; Wang, M.; Qian, B.; Yang, B.; Li, Z.; Shi, L. X. Ensemble just-in-time learning-based soft sensor for mooney viscosity prediction in an industrial rubber mixing process. *Adv. Polym. Technol.* **2020**, *2020*, No. 6575326.
(30) Guo, F.; Bai, W. T.; Huang, B. Output-relevant variational autoencoder for just-in-time soft sensor modeling with missing data. *J. Process Control* **2020**, *92*, 90−97.
(31) Huang, G.; Huang, G.; Song, S.; You, K. Trends in extreme learning machines: A review. *Neural Networks* **2015**, *61*, 32−48.
(32) Houthuys, L.; Langone, R.; Suykens, J. A. K. Multi-view least squares support vector machines classification. *Neurocomputing* **2018**, *282*, 78−88.
(33) Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; WardeFarley, D.; Ozair, S.; Courville, A.; Bengio, Y. In *Generative Adversarial Nets*, 28th Conference on Neural Information Processing Systems, Montreal, Canada, December 8−13, 2014.
(34) Nahas, E. P.; Henson, M. A.; Seborg, D. E. Nonlinear internal model control strategy for neural network models. *Comput. Chem. Eng.* **1992**, *16*, 1039−1057.

https://doi.org/10.1021/acsomega.3c01832
ACS Omega 2023, 8, 19900−19911