Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-18 (2018) 608–613

PET Viscosity Prediction Using JIT-based Extreme Learning Machine

Zhenxing Li[1], Kuangrong Hao[1,2*], Lei Chen[1,2], Yongsheng Ding[1,2], Biao Huang[3]

1. College of Information Science and Technology, Donghua University, Shanghai 201620 China (e-mail: krhao@dhu.edu.cn)*

2. Engineering Research Center of Digitized Textile & Apparel Technology, Ministry of Education, Shanghai 201620 China

3. Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada

Abstract: As a key stage in polyester production, polymerization process is difficult to model due to its complex reaction mechanism. As a result, online viscosity prediction in industrial polyester polymerization processes is not an easy task. An efficient data-driven prediction model is considered in this work. In order to solve the problem of low accuracy of the online viscosity measuring instrument and considerably time-consuming laboratory analysis, variables that are easily monitored during the polymerization process, i.e. temperature and pressure in the main reactor as well as the viscometer values, are selected to establish an Extreme Learning Machine (ELM) viscosity prediction model. A Just-in-time-based ELM model was established to predict the viscosity values under multi-mode operating and multi-standard production conditions. Consequently, without relying on the time-consuming laboratory analysis process, the PET viscosity can be predicted online. The industrial PET viscosity prediction results show the improved prediction performance of the proposed modeling approach in comparison with ELM and JPCR (Just-in-time principal component regression) approaches.

Keywords: Polyethylene terephthalate; polyester polymerization processes; PET viscosity prediction; Extreme Learning Machine; Just-in-time Learning.

1. INTRODUCTION

PET (Polyethylene terephthalate) is the first polyester product to be industrialized and widely used in the world, which is polycondensed with PTA (Terephthalic acid) and EG (Ethylene glycol) as the raw material. It is also the largest and the most widely used polymer synthetic material. Due to the increasingly complex, non-linear and time-varying characteristics of the process, data-driven modelling has become a favourable approach. The performance of the model for product polyester polymerization process is particularly important for the production. As one of the important indices of PET quality, the higher viscosity means the longer of the molecular chain. The stronger the interaction between molecules is, the stronger the mechanical strength is. Polyester resins with different viscosities are used to make different products for different purposes due to their different mechanical properties. Therefore, viscosity prediction is becoming increasingly important in the polyester polymerization process. However, since performance of most online viscosity measuring instruments decreases with increasing time and the development of mechanistic models for viscosity prediction is time consuming and laborious, Gonzaga et al. (2009) built a soft-sensor based on a feed-forward artificial neural network to provide on-line estimates of the PET viscosity. Cao et al. (2006) proposed a Hybrid Neural Networks to predict the viscosity of the PET well. Many researchers have used neural networks to predict other quality indicators of polyester (Barton and Himmelblau, 1997; Ponnuswamy et al., 2010; Tian et al., 2010; Zhang et al., 1998).

In recent years, data driven and artificial intelligence technologies have been applied widely in systems modelling and chemical process control, and have demonstrated better performance than traditional methods. Artificial neural networks have also been used in complex industrial modelling because of their adaptive and self-learning ability. However, traditional neural networks often exhibit disadvantages with their complex structure, slow learning speeds and long training time. proposed a novel neural network structure of a single hidden layer, which is based on Extreme Learning Machine (ELM). In the ELM, the input layer weight and hidden layer threshold value are generated randomly, and the output weight are determined by analysis. ELM and its variants are widely used in data processing(Liang et al., 2006), unsupervised learning(G. Huang et al., 2017; Zong et al., 2013), biomedicine(Song and Zhang, 2013; You et al., 2013), computer vision, image processing(An and Bhanu, 2013), system modelling and forecasting(Du et al., 2013) owing to their fast convergence and generalization ability. Liu et al. (2012) used the near infrared spectroscopy system to obtain the near-infrared spectra from the solid-state fermentation samples. The reference values of the near-infrared spectroscopy were analysed experimentally and a prediction model of pH was built using the ELM algorithm. Jin et al. (2017) established a JRELM model in industrial rubber mixing processes to make an on-line prediction of Mooney viscosity.

However, artificial neural networks and other nonlinear models were less successful for multi-mode and multi-specification production processes along with large-scale data. Just-in-time learning (JITL)(Cybenko, 1996) has been a

10.1016/j.ifacol.2018.09.357

commonly used modelling strategy for nonlinear processes. It is originated from database technology and local modelling techniques, also known as lazy learning(Bontempi et al., 2001), locally weighted learning(Christopher et al., 1997), instance-based learning(Aha et al., 1991) and model-on-demand(Braun et al., 2001). The global modelling and the traditional local modelling techniques cannot build an effective model when dealing with large-scale strong nonlinear processes. JITL modelling algorithm can select the data that is similar as the query data from the historical dataset to build a local model. This allows for better modelling accuracy in tracking quick process changes and handling strong nonlinearity. In recent years, JITL technology has been applied to industrial data that is time-varying with large random noise(Yuan et al., 2016), and process that is nonlinear. Kim et al. (2013) developed a control system that integrates MPC software and JITL-based soft sensors for applications in the cracked gasoline fractionation tower, which reduces operating costs by approximately 0.6% and significantly reduces environmental impacts. Based on JITL and ELM, Peng et al. (2017) has proposed an on-line monitoring and modelling method for performance prediction and pattern recognition of non-Gaussian multimode processes.

The main contributions of this paper are: 1) In order to solve the problem of the inaccurate online viscosity measuring instrument and the time-consuming of laboratory analysis, the data that are easily monitored during the polymerization process, i.e. temperature and pressure in the main reactor and viscometer value, are selected to establish an Extreme Learning Machine (ELM) viscosity prediction model. 2) A Just-in-time ELM (JIT-ELM) model was established for the multi-mode and multi-standard product viscosity. Consequently, without the time-consuming laboratory analysis process, the PET viscosity can be predicted online. The industrial PET viscosity prediction results show improved prediction performance in comparison with ELM and JPCR approaches.

The remainder of this paper is organized as follows: In Section 2, we elaborate the mathematical formulation for ELM and JIT-ELM algorithm. In Section 3, the polyester polymerization processes are discussed in detail. Section 4 describes the application of JIT-ELM for viscosity prediction. Conclusions are drawn in Section 5.

2. THE ELM MODEL AND JUST-IN-TIME LERANING ALGORITHM

2.1. Extreme Learning Machines

Single-hidden layer feed-forward neural network (SLFN) has been widely used in many fields owing to its good learning ability(G. B. Huang, et al., 2006). However, some of the shortcomings inherent in traditional learning algorithms are the main bottlenecks that constrain their applications. Most of the feed-forward neural networks use the gradient descent method, which has the following shortcomings: (1) since the gradient descent approach requires multiple iterations to modify the weights and thresholds, the training process is slow. (2) It is easy to fall into the local minimum. (3) Learning rate η has a greater impact on the performance of neural networks, and it is difficult to choose the appropriate η. If η is too small, the convergence rate of the algorithm is slow. On the other hand, if η is too large, the training process can be unstable and potentially affect the convergence.

Traditional SLFN is a fully connected network which has the input layer, the hidden layer and the output layer. There are n neurons in the input layer, corresponding to n input variables. The hidden layer has l neurons. The output layer has m neurons, corresponding to m output variables. The corresponding output of SLFN can be expressed as:

$$f(\mathbf{x}_i) = \sum_{j=1}^{l} \boldsymbol{\beta}_j g_j(\mathbf{x}_i) = \sum_{j=1}^{l} \boldsymbol{\beta}_j g(\boldsymbol{w}_j, \boldsymbol{b}_j, \mathbf{x}_i), i = 1, 2, \cdots, Q \quad (1)$$

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ln} \end{bmatrix}_{l\times n} \quad \beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_{l1} & \beta_{l2} & \cdots & \beta_{lm} \end{bmatrix}_{l\times m} \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_l \end{bmatrix}_{l\times 1} \quad (2)$$

where w_{ji} represents the connection weight between the ith neuron of the input layer and the jth neuron of the hidden layer; β_{jk} represents the connection weight between the jth neuron of the hidden layer and the kth neuron of the output layer; b is the threshold of the hidden layer; $g(x)$ is the activation function of the neurons of the hidden layer.

Suppose there are Q samples in the training input matrix X and the output matrix Y:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ x_{21} & x_{22} & \cdots & x_{2Q} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{bmatrix}_{n\times Q} \quad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mQ} \end{bmatrix}_{m\times Q} \quad (3)$$

The output of the network is $T = \left[t_1, t_2, \cdots, t_Q\right]_{m\times Q}$, where

$$t_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m\times 1} = \begin{bmatrix} \sum_{i=1}^{l} \beta_{i1} g(w_i x_i + b_i) \\ \sum_{i=1}^{l} \beta_{i2} g(w_i x_i + b_i) \\ \vdots \\ \sum_{i=1}^{l} \beta_{im} g(w_i x_i + b_i) \end{bmatrix}_{m\times 1}, \; j = 1, 2, \cdots, Q \quad (4)$$

Equation 4 can be further expressed as:

$$\boldsymbol{H}\beta = T' \quad (5)$$

Where T' is the transpose of the matrix T; $\boldsymbol{H}$ is the hidden layer output matrix of the neural network. The output of the network can be obtained after calculating the value of β. Consequently, given a data set $N = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n}$, hidden layer neurons l and activation function $g(x)$, the ELM algorithm steps are summarized as follows:

(1) Randomly generate input weights and hidden layer threshold weights $(\boldsymbol{w}_j, \boldsymbol{b}_j)$;

(2) Calculate the hidden layer output matrix $\boldsymbol{H}$;

(3) Calculate the output weight vector $\hat{\boldsymbol{\beta}} = \boldsymbol{H}^{\dagger}\boldsymbol{T}'$.

(4) With $\hat{\boldsymbol{\beta}}$, the ELM model can then be established.

2.2. Just-in-time Learning Algorithm

Compared to other methods, just-in-time learning has the following characteristics: First, only when there is need to query the output of a given input sample, the modelling process is activated. Thus, the method is an on-line modelling technique. Second, the output prediction uses only the partial samples of the historical data that are most relevant to the query sample and not all historical samples are used. Third, once the prediction of the query sample is completed, the established local model is discarded until the next query sample arrives (Cheng and Chiu, 2004). Specifically, real-time learning modelling is divided into three steps:

(1) Relevant sample selection

Relevant sample selection is the primary feature of just-in-time learning and is the most critical reason for just-in-time learning to achieve good prediction performance. Sample selection is often based on the distance between the historical samples and query samples. Common distance calculation is based on Euclidean distance and angle distance (Cheng and Chiu, 2004):

$$d_i = \left\| \mathbf{x}_q - \mathbf{x}_i \right\|_2 \tag{7}$$

$$\cos(\theta_i) = \frac{\mathbf{x}_q^{\mathrm{T}} \mathbf{x}_i}{\left\| \mathbf{x}_q \right\|_2 \cdot \left\| \mathbf{x}_i \right\|_2} \tag{8}$$

If $\cos(\theta_i) \geq 0$, the similarity s_i is calculated as:

$$s_i = \gamma \sqrt{e^{-d_i^2}} + (1-\gamma)\cos(\theta_i) \tag{9}$$

If $\cos(\theta_i) < 0$, the data $(\mathbf{x}_i, \mathrm{y}_i)$ is discarded.

Here, $\mathbf{x}_i$ represents the *ith* historical sample, and $\mathbf{x}_q$ represents the query sample at which the output is to be predicted. In general, the greater the distance between the historical sample and the query sample, the lower the similarity between the historical sample and the query sample, so that the likelihood that the corresponding data will be selected for modelling is lower. After the distance calculation, the weight of each sample value must be specified. The weighting function is generally designed as a function of distance. As the distance increases, the similarity between the historical sample and the query sample is reduced, and the weight of the corresponding historical sample is smaller.

(2) Local modelling

Relevant historical samples for local modelling can be selected after the distance and weight calculations are completed. Assume that the input and output of the selected historical samples are $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$, respectively. Local Weighted Regression (LWR) is the most common modelling algorithm in just-in-time learning. LWR can be fitted to any smooth non-linear function through piecewise linear relationship. However, in the local modelling phase, there is no restriction on what type of the regression algorithms can be used.

(3) Predict output

Finally, when the local model is established, the output of the query sample can be predicted. Once the output prediction is completed, the model is discarded. When the next query sample arrives, JIT starts over again.

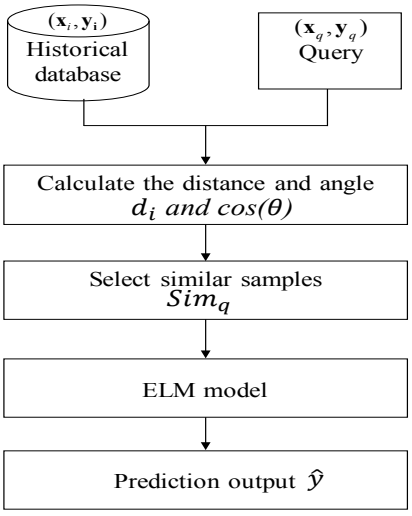

Fig. 1. The flow chart of JIT-ELM model

2.3. The JIT-ELM Algorithm

As mentioned above, the LWR can fit to any smooth non-linear function through piecewise linear relationship in the local modelling stage. However, in many industrial applications, data often exhibit high dimensionality and strong correlations. LWR algorithm is prone to problems such as matrix morbidity and algorithm instability. Most chemical processes have complex chemical reactions, and polyester polymerization process is of no exception. This leads to the prevalence of non-linear relationship between the variables. Therefore, this paper applies the ELM for local modelling owing to the advantages of fast learning speed and good generalization performance. The flow chart of the adopted modelling process is shown in Fig. 1. The JIT-ELM algorithm has the following steps:

JIT-ELM Step 1: Build the historical dataset $(\mathbf{x}_i, \mathbf{y}_i)$ and the query sample set $(\mathbf{x}_q, \mathbf{y}_q)$;

JIT-ELM Step 2: When a new query sample arrives, using (8), (9), and (10) to compute degrees of similarity in historical data sets and create similar datasets Sim_q;

JIT-ELM Step 3: List Sim_q in descending order and select $\boldsymbol{p}$ most relevant data;

JIT-ELM Step 4: Use the $\boldsymbol{p}$ data to train the ELM network to get the output for the query data $(\mathbf{x}_q, \mathbf{y}_q)$;

JIT-ELM Step 5: Once the predicted value is obtained, the model is discarded. When there is a new query data, repeat steps 2, 3, and 4.

When the new query data arrives, the above five steps can be implemented to establish the model for the forecast of output.

3. THE PET POLYMERIZATION PROCESS

Compared to other chemical reactions, polymerization and polymer production have the following two characteristics: (1) the reaction mechanism has a diverse, complex dynamics, poor reproducibility, and trace impurities. In addition to the polymerization rate, the molecular weight and its distribution, copolymer composition and sequence distribution, polymer structure and polymer properties need to be considered in the polymerization process modeling. (2) The polymerization reaction is of high viscosity, and results in non-Newtonian fluids or high solid content of the suspension or emulsion, and some gas or gas-liquid-solid multiphase system. It is difficult to measure the chemical basis data (especially rheological properties) of these systems, and it is more difficult to further study the transmission process such as flow, stirring, mixing, heat transfer and mass transfer. Additionally, it is of not only concern about the balance of materials and heat balance, but it is of more concern about the polymer itself, such as the average molecular weight, molecular weight distribution, polymer activity chain, etc. in the modelling of polymerization process. The difficulty in making a qualitative analysis and then giving a quantitative description to establish a mathematical model for the above aspects of the complex problems is self-evident. Consequently, the prediction of the polyester product quality is difficult.

The case study is based on a process from ZHONGXIN Chemical Fiber Co., Ltd. polyester plant located in TongXiang, Zhejiang, China. The factory has a total of 10 production lines. This study only selected one with the daily production capacity of 500 tons of pre-oriented yarn (POY). The factory uses DuPont PET production technology, and the PTA process flow chart is shown in Fig. 2.

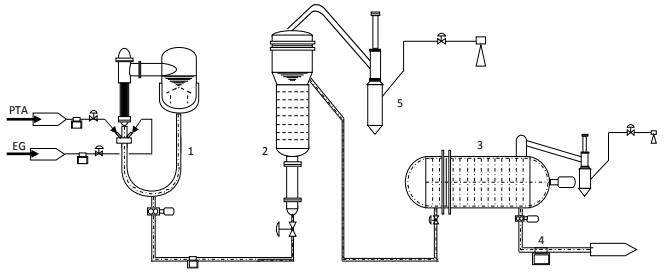

Fig. 2. DuPont PTA process flow chart

The main device in Fig. 2 is shown in Table 1. DuPont craft is a common polyester industrial manufacturing method. It uses three reactors, the esterification reactor, a tower-type upstream tank reactor and a horizontal squirrel cage stirred tank reactor, along with EG sprinkler system for the room temperature jet condenser and the vacuum power system with a high pressure steam jet pump. This paper only studies the prediction of the polymerization process.

Table 1. The main device in Fig. 2

Number	Device Name
1	Esterification reactor (ER)
2	Pre-polycondensation reactor(PR)
3	Final polycondensation reactor(FP)
4	Viscometer
5	Condenser

4. JIT-ELM MODEL APPLICATION TO INDUSTRIAL VISCOSITY PREDICTION: RESULTS AND DISCUSSION

With the rapid development of computer technology and the extensive application of distributed control system (DCS) in the industry, the production enterprises generate and store a large amount of process data every day. In this work we collected 38 days of factory production data from DCS, a total of 55,322 data points. In order to ensure that the data can truly reflect the actual factory production, a preliminary set of input variables for the JIT-ELM model was selected according to a priori knowledge about the process. These input variables were further analysed considering their correlations with the polymer viscosity by using sensitivity analysis(Gonzaga, et al., 2009). Based on these analyses, five variables were selected at the input of the JIT-ELM model. The output references used in the output layer of the JIT-ELM (the target variable) during the training stage were the actual viscosity values obtained from the plant viscometer data records. The input and output variables of the JIT-ELM model are shown in Table 2. To build a model under multiple conditions and multi-standard production, 1000 data points were randomly selected as the query (test) samples from all 55,322 data, and the remaining 54,322 data was used as historical data sets for training.

Table 2. The input data and output data of the JIT-ELM model

Input	T-1	°C	The temperature of the ER
	T-2	°C	The temperature of the PR
	T-3	°C	The temperature of the FP
	P-1	mmHg	The pressure of the PR
	P-2	mmHg	The pressure of the FP
Output	V-1	CU	The value of the viscometer

Three modelling methods, including JIT-ELM, ELM, and JIT-based PCR (simply denoted as JPCR), are compared in this work. Fig. 3, 4, 5 show the results of JIT-ELM, JPCR and ELM for the online prediction of viscosity, respectively. From the

distribution of the predictions in the graph, it can be seen that the JIT-ELM model outperforms the other two models. JPCR and ELM have poorer predictions. Comparing between JPCR and ELM, JPCR shows better predictions than ELM. Fig. 6 shows the box plot of the predicted and true values. It can be clearly seen from the box plot that the accuracy of the results of the proposed algorithm is indeed higher than that of the JPCR and ELM algorithms.

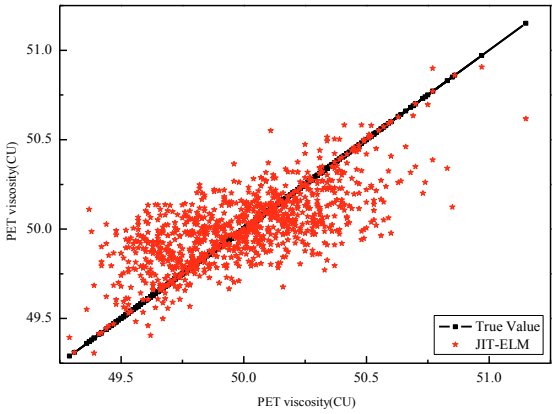

Fig. 3. The proposed algorithm

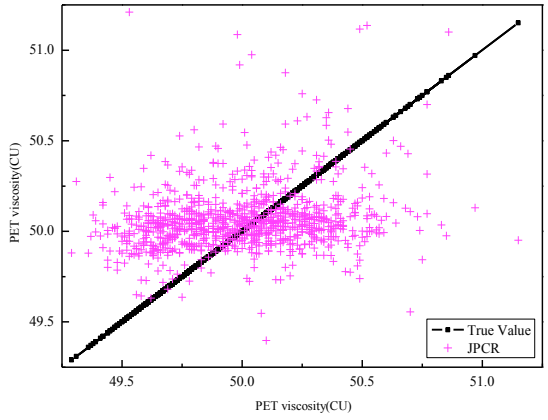

Fig. 4. The JPCR algorithm

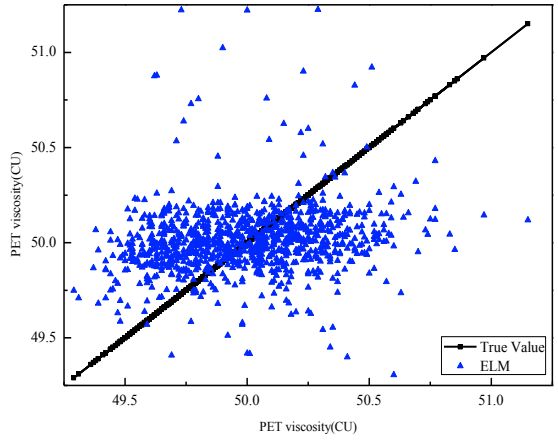

Fig. 5. The ELM algorithm

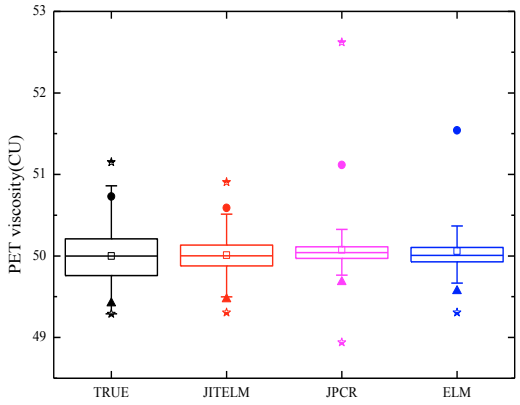

Fig. 6. The box plot of the predicted and true values.

The performance of JIT-ELM model is verified by means of MSE and RMSE, defined below:

Mean square error (MSE):

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y})^2 \tag{10}$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y})^2} \tag{11}$$

Where y_i and $\hat{y}_i$ are the true value and the prediction value, respectively, and m is the number of test samples.

The comparisons are shown in Table 3. It is clear that both RMSE and MSE of JIT-ELM are smaller than the other two models, indicating that the proposed algorithm produces more accurate predictions.

Table 3. RMSE and MSE comparison of the three models

Algorithm	RMSE	MSE
ELM	0.4306	0.1854
JPCR	0.3524	0.1242
JIT-ELM	**0.2095**	**0.0439**

5. CONCLUSIONS

This paper adopts a combined Extreme Learning Machine and Just-in-time learning approach for viscosity prediction in PET production. According to the process understanding of the polymerization process, we select the temperature of ER, PR and FP and the pressure of PR and FP as the input variables, and the viscometer measurements as the reference data. Error analysis and comparative studies of the prediction results demonstrate the effectiveness of the developed model. A good prediction of viscosity is of practical importance for high-performance fibre products with stable and excellent performance.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Plan from Ministry of Science and

Technology (2016YFB0302701), National Natural Science Foundation of China (nos. 61503075, 61473077, 61473078,), International Collaborative Project of the Shanghai Committee of Science and Technology (no. 16510711100), and Program for Changjiang Scholars from the Ministry of Education (2015-2019).

REFERENCES

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning, 6*, 37-66.

An, L., & Bhanu, B. (2013). Image super-resolution by extreme learning machine. In *IEEE International Conference on Image Processing* (pp. 2209-2212).

Barton, R. S., & Himmelblau, D. M. (1997). Online prediction of polymer product quality in an industrial reactor using recurrent neural networks. In *International Conference on Neural Networks* (pp. 111-114 vol.111).

Bontempi, G., Bersini, H., & Birattari, M. (2001). The local paradigm for modeling and control: from neuro-fuzzy to lazy learning. *Fuzzy sets and systems, 121*, 59-72.

Braun, M. W., Rivera, D. E., & Stenman, A. (2001). A'Model-on-Demand'identification methodology for non-linear process systems. *International Journal of Control, 74*, 1708-1717.

Cao, L. L., Xu, X. H., & Jiang, P. (2006). Viscosity Prediction for PET Process Based on Hybrid Neural Networks. In *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on* (pp. 4871-4875).

Cheng, C., & Chiu, M.-S. (2004). A new data-based methodology for nonlinear process modeling. *Chemical engineering science, 59*, 2801-2810.

Christopher, A., Andrew, M., & Stefan, S. (1997). Locally weighted learning. *Artif Intell Rev, 11*, 11-73.

Cybenko, G. (1996). Just-in-Time Learning and Estimation. *Nato Asi*.

Du, D., Li, K., Irwin, G. W., & Deng, J. (2013). A novel automatic two-stage locally regularized classifier construction method using the extreme learning machine. *Neurocomputing, 102*, 10-22.

Gonzaga, J. C. B., Meleiro, L. A. C., Kiang, C., & Filho, R. M. (2009). ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering, 33*, 43-49.

Huang, G., Song, S., Gupta, J. N., & Wu, C. (2017). Semi-supervised and unsupervised extreme learning machines. *IEEE Transactions on Cybernetics, 44*, 2405-2417.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*, 489-501.

Jin, W., Liu, Y., & Gao, Z. (2017). Fast property prediction in an industrial rubber mixing process with local ELM model. *Journal of Applied Polymer Science, 134*, n/a-n/a.

Kim, S., Kano, M., Hasebe, S., Takinami, A., & Seki, T. (2013). Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges. *Industrial & Engineering Chemistry Research, 52*, 12346-12356.

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Transactions on Neural Networks, 17*, 1411-1423.

Liu, G. H., Jiang, H., Xiao, X. H., Zhang, D. J., Mei, C. L., & Ding, Y. H. (2012). Determination of process variable pH in solid-state fermentation by FT-NIR spectroscopy and extreme learning machine (ELM). *Spectroscopy & Spectral Analysis, 32*, 970.

Peng, X., Tang, Y., Du, W., & Qian, F. (2017). An Online Performance Monitoring and Modeling Paradigm based on Just-in-time Learning and Extreme Learning Machine for Non-Gaussian Chemical Process. *Industrial & Engineering Chemistry Research, 56*.

Ponnuswamy, S., Shah, S. L., & Kiparissides, C. (2010). On-line monitoring of polymer quality in a batch polymerization reactor. *Journal of Applied Polymer Science, 32*, 3239-3253.

Song, Y., & Zhang, J. (2013). Automatic recognition of epileptic EEG patterns via Extreme Learning Machine and multiresolution feature extraction. *Expert Systems with Applications, 40*, 5477-5489.

Tian, Y., Zhang, J., & Morris, J. (2010). Modeling and Optimal Control of a Batch Polymerization Reactor Using a Hybrid Stacked Recurrent Neural Network Model. *Industrial & Engineering Chemistry Research, 40*, 4525-4535.

You, Z. H., Lei, Y. K., Zhu, L., Xia, J., & Wang, B. (2013). Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC bioinformatics, 14*, 1-11.

Yuan, X., Huang, B., Ge, Z., & Song, Z. (2016). Double locally weighted principal component regression for soft sensor with sample selection under supervised latent structure. *Chemometrics and Intelligent Laboratory Systems, 153*, 116-125.

Zhang, J., Morris, A. J., Martin, E. B., & Kiparissides, C. (1998). Prediction of polymer quality in batch polymerisation reactors using robust neural networks. *Chemical engineering journal, 69*, 135-143.

Zong, W., Huang, G. B., & Chen, Y. (2013). Weighted extreme learning machine for imbalance learning. *Neurocomputing, 101*, 229-242.