

Prediction of Physical Properties of Crude Oil Based on Ensemble Random Weights Neural Network

Jun Lu *. Jinliang Ding*. Changxin Liu*. Yaochu Jin*. **.

*State Key Laboratory of Synthetical Automation for Process Industry, Northeastern University, Shenyang, 110819, China (e-mail: junluchn@hotmail.com, {jlding, cxliu }@mail.neu.edu.cn).

** Department of Computer Science, University of Surrey, Guildford, Surrey, United Kingdom (e-mail: yaochu.jin@surrey.ac.uk)

Abstract: Prediction of physical properties of crude oil plays a key role in the petroleum refining industry, therefore, it is of great significance to establish the prediction model of physical properties of crude oil. In this paper, we propose an ensemble random weights neural network based prediction model whose inputs are nuclear magnetic resonance (NMR) spectra and outputs are carbon residual and asphaltene of crude oil. The model uses random vector functional link (RVFL) networks as the basic components and employs the regularized negative correlation learning strategy to build neural network ensemble and the online method to learn the new data. The experiment using the practical data collected from a refinery is carried out and compared with the decorrelated neural network ensembles with random weights (DNNE), least squares support vector machine (LS-SVM), partial least squares regression (PLS) and multiple linear regression (MLR). The results indicate the effectiveness of the proposed approach.

Keywords: physical properties of crude oil, prediction model, ensemble random weights, neural network; online learning.

1. INTRODUCTION

With the rapid development of the world economy, oil refineries and petrochemical companies in China are confronted with increasingly competitive pressure from the international oil companies (Baruník and Malinská 2015). In order to reduce the production costs, online blending technology for crude oil is adopted by refineries to make reasonable mixing of various kinds of crude oil. Prediction of physical properties of crude oil is the precondition of online blending of crude oil, therefore fast evaluation of physical properties of crude oil before refining is important to refineries (Sriram et al. 2007).

Carbon residual and asphaltene of crude oil are two significant physical properties in crude oil blending. The value of carbon residual is the percentage of the residue carbon formed in the oil after the evaporation and pyrolysis process at the especially high temperature condition. The value of carbon residual can be used to characterize the relative coke tendency of oil, which is regarded as one of the selection criteria of raw material and production process, and it is also one of the important control indexes in the production process. The value of carbon residual in the catalytic cracking material is an important parameter to evaluate the crude oil. The deposition of asphaltenes in the petroleum industry can result in a significant increase of production costs. Asphaltenes in crude oil contain higher acid functional groups and a large amount of condensed aromatic hydrocarbons that seriously affect the refining process and lead to the blockage of pipelines. Thus, prediction of the two physical properties of crude oil is extremely important.

Most of the fast evaluation systems of physical properties of crude oil are based on near infrared (NIR) technology at present (He et al. 2015). Higher quality of crude oil and longer optical path are required by NIR, but its sensitivity is relatively low. With the development of computer technology, nuclear magnetic resonance (NMR) has become one of the most advanced and promising process analysis techniques. The efficiency of industrial production is improved significantly by applying NMR combined with the advanced control technology. In some international oil companies such as BP, BASF, Shell and AGIP, NMR is used for monitoring and control in the crude oil distillation and blending, gasoline and diesel blending, catalytic cracking, catalytic reforming and ethylene cracking processes (Lin et al. 2004). Online measured data are used to supervise the operation process which can improve the oil yield and quality that bring great economic benefits to the petrochemical industry. NMR requires simple sample pretreatment and the accuracy is not affected by viscosity, purity, color and water content of the sample. However, the prediction models of physical properties of crude oil based on NMR are inadequate at present as most are linear models such as partial least squares (PLS) regression (Behera et al. 2008) and multiple linear regression (MLR) (Barbosa et al. 2016) which are impossible to characterize the complicated relation between the NMR spectrum and the physical properties of crude oil accurately. Effective prediction model is the premise and basis of fast evaluation and analysis of crude oil and it is an urgent problem to be solved.

A prediction model of carbon residue and asphaltene of crude oil based on ensemble random weights neural network is put

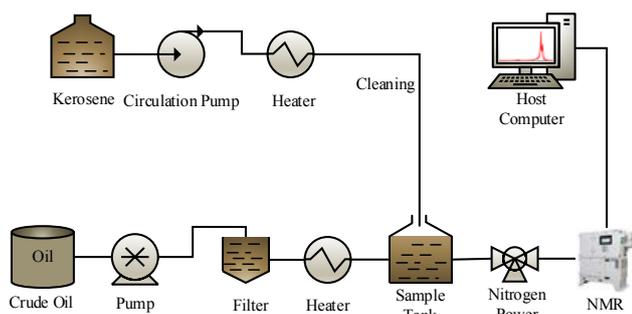


Fig.1. Structure chart of fast evaluation system of crude oil based on NMR.

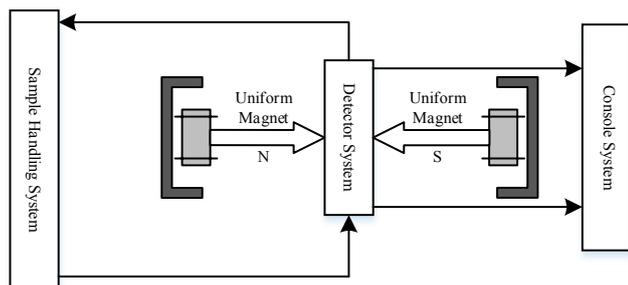


Fig.2. Schematic diagram of NMR.

forward in this paper to solve the problem that it is difficult to establish the mechanism model between NMR spectrum and the physical properties of crude oil. The model can be used to obtain evaluation data of crude oil which can provide fast, simple, effective and available information for the determination of crude oil processing scheme and the optimization of production decision. The proposed model also lays foundation for the application of NMR technology in petrochemical industry. The algorithm proposed in this paper can be extended to the prediction and analysis of other physical properties of crude oil which has extraordinary broad application prospects.

The organization of the rest of this paper is as follows. In section II, the evaluation system of physical properties of crude oil based on NMR is presented. The prediction model is given in section III, where the ensemble random weights neural network is discussed in detail. In section IV, the experiment is carried out and the results are discussed. At last, our conclusion is given in section V.

2. EVALUATION SYSTEM OF PHYSICAL PROPERTIES OF CRUDE OIL BASED ON NUCLEAR MAGNETIC RESONANCE

The fast evaluation system of crude oil based on nuclear magnetic resonance mainly consists of filter, heater, nuclear magnetic resonance and host computer. First, the crude oil sample is sent to the sample tank after filtration and heating. Then the sample is taken to the NMR by nitrogen power and the results are uploaded to the host computer by the communication system. Finally, physical properties of crude oil are obtained by the NMR spectrum and the prediction model. Fig.1 is the structure chart of fast evaluation system of crude oil based on NMR.

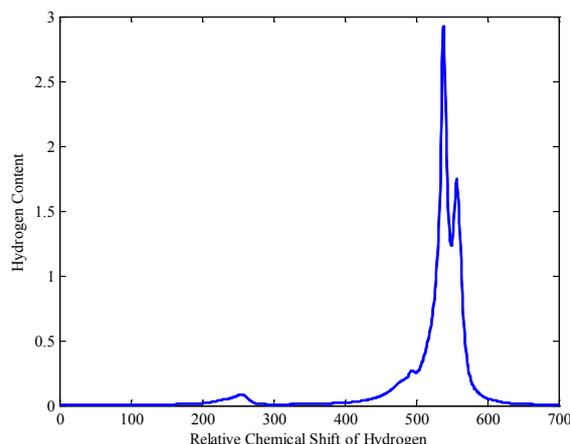


Fig.3. Typical NMR spectrum.

When the NMR is used to measure the sample, the sample passes through a precisely controlled magnetic field and all the protons in the sample are aligning in the direction of the magnetic field. NMR emits the pulse electromagnetic energy to the sample and then the directions of the alignment of protons will shift. When the radio frequency signal is removed, the protons release energy and eventually return to their original positions. During this process, the protons will produce the free induction decay (FID) signals, then the analog signals are handled by fast Fourier transform (FFT) and converted to a NMR spectrum. All the steps mentioned above could be completed automatically. Fig.2 is the schematic diagram of NMR.

The data used in this paper are collected from a refinery in south China. The spectrum is analysed by Aspect Imaging AI-60 online NMR. Fig.3 is the spectrum obtained at the conditions that the operating frequency is 60MHz and the chemical shift reference is tetramethylsilane (TMS).

3. PREDICTION MODEL OF PHYSICAL PROPERTIES OF CRUDE OIL BASED ON ENSEMBLE RANDOM WEIGHTS NEURAL NETWORK

From the analysis above, we know that there is a complex relation between the NMR spectrum and the physical properties of crude oil. While most of the prediction models are linear models such as PLS and MLR at present, which are impossible to characterize the complex relation between the NMR spectrum and the physical properties of crude oil. Thus, the complicated data-based models should be selected because of their abilities to characterize complex relationships.

3.1 Selection of Model Input and Output

According to the evaluation system of physical properties of crude oil based on NMR and the modelling objectives, the input vector of the prediction model is selected as the NMR spectrum data $x \in R^{700}$. The output of the prediction model is the physical property of crude oil $y \in R^1$. In this paper, the prediction models of carbon residual content and asphaltene content which are two important physical properties for evaluating the quality of crude oil are built.

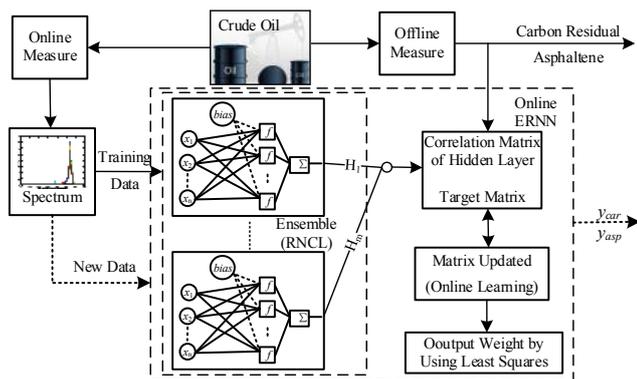


Fig.4. Structure chart of prediction model.

3.2 Structure of the Prediction Model

Artificial neural network based on back-propagation (BP) algorithm has been used to predict the properties of crude oil using NMR (Shea and SeldaGunsel. 2003). But the BP algorithm is easy to suffer from local optimum. To address this issue, neural network which has random input weights and biases of hidden nodes is presented (Igelnik and Pao. 1995). But there is no unified theory to guide the selection of the structure and parameter of neural network, so researchers' experience may determine its performance. In order to solve the problem, neural networks ensemble is proposed (Hansen and Salamon. 1990). While the ensemble model may result in over fitting. To solve the issue, a regularized negative correlation learning (RNCL) is proposed (Chen and Yao. 2009). In this paper, the ensemble random weights neural network based on regularized negative correlation learning is adopted to establish the prediction model. The structure chart of the proposed model is illustrated in Fig.4 and the details are discussed in the following.

3.3 Ensemble Random Weights Neural Network Based Prediction Model

The existing ensemble techniques are simple combinations of all the individual models and only minimize the empirical error, hardly take the diversity and complexity of the model into account (Brown et al. 2005; Chong et al. 2009; Wang and Yao. 2012). To solve this problem, an ensemble neural network with random weights (ERNN) based prediction model is adopted in this paper.

Neural Network Ensemble based on Negative Correlation Learning: Considering a given sample $\{x(k), y(k)\}$, where $x(k) \in R^{700}$ is NMR spectrum and $y(k)$ is the physical property of crude oil. The RVFL whose input weights and hidden layer biases are generated randomly (Igelnik and Pao. 1995) can be written as:

$$f(x(k)) = \sum_{i=1}^L \beta_i G(w_i^T \cdot x(k) + b_i) = \sum_{i=1}^L \beta_i H_i \quad (1)$$

where β_i denotes the output weight of i th node, H_i is the output of i th hidden node, L represents the number of hidden

nodes, w_i and b_i are input weights and hidden layer bias of i th hidden node and $G(\cdot)$ is the activation function.

The output of ensemble neural network (Chen and Yao. 2009) based prediction model is the linear combinations of RVFLs:

$$f_{ens}(x(k)) = \omega_m \sum_{m=1}^M f_m(x(k)) \quad (2)$$

where $f_m(x(k))$ is the output of the m th basic RVFL, ω_m is the weight of the m th basic RVFL and in this paper, ω_m is set as $1/M$.

However, the correlations between the basic models are not taken into account in the ensemble learning above. To address this issue, negative correlation learning (NCL) is proposed (Liu and Yao. 1999). In the NCL ensemble neural network model, the cost function e_m of each basic model $f_m(x(k))$ is defined as follows:

$$e_m = \frac{1}{2} \sum_{k=1}^N (f_m(x(k)) - y(k))^2 + \lambda p_m \quad (3)$$

where p_m is the correlation penalty function, λ is penalty function parameter and $\lambda \in [0, 1]$.

$$p_m = -\sum_{k=1}^N (f_m(x(k)) - f_{ens}(x(k)))^2 \quad (4)$$

Regularized NCL based Neural Network Ensemble: By introducing a regularization function, the RNCL neural network is established. The cost function of each individual network in RNCL (Chen and Yao. 2009) can be shown as:

$$\begin{aligned} e_m &= \frac{1}{2} \sum_{k=1}^N (f_m(x(k)) - y(k))^2 + \lambda p_m + \alpha_m \|\beta_m\|_2^2 \\ &= \frac{1}{2} \|H_m \beta_m - T\| + \alpha_m \|\beta_m\|_2^2 - \lambda \left\| H_m \beta_m - \frac{1}{M} \sum_{j=1}^M H_j \beta_j \right\|_2^2 \end{aligned} \quad (5)$$

where H_m is the hidden layer output matrix of the basic model m , $T = [y(1), y(2), \dots, y(N)]^T$ is the target matrix, $\beta_m = [\beta_{m1}, \beta_{m2}, \dots, \beta_{mL}]^T$ is the output weights of the basic model m and $\alpha_m \in [0, 1]$ is the regularization parameter.

When the gradient of e_m vanishes with respect to the output weight β_m , the RNCL neural network ensemble can get a best generalization performance. It can be given as:

$$(C_1 H_m^T H_m + 2\alpha_m I) \beta_m + C_2 \sum_{\substack{j=1 \\ j \neq m}}^M H_m^T H_j \beta_j = H_m^T T \quad (6)$$

where C_1 and C_2 are two constant parameters which can be expressed as follows:

$$C_1 = 1 - 2 \times \lambda \left(1 - \frac{1}{M}\right) + 2 \times \frac{\lambda}{M} \left(1 - \frac{1}{M}\right) \quad (7)$$

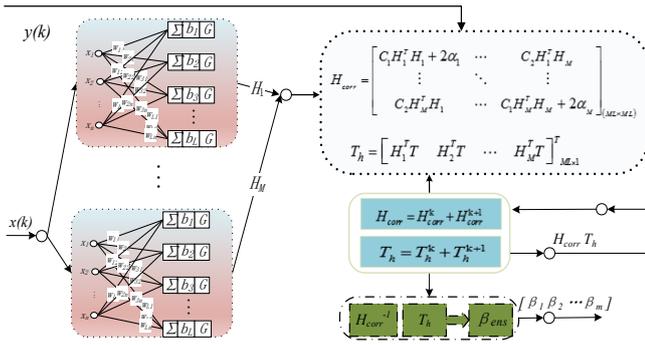


Fig.5. Structure chart of online learning ERNN.

$$C_2 = 2 \times \frac{\lambda}{M} \left(1 - \frac{1}{M}\right) \quad (8)$$

The relationship between the hidden layer and the output layer can be denoted in a matrix form as follows:

$$H_{corr} \beta_{ens} = T_h \quad (9)$$

$$T_h = [H_1^T T, H_2^T T, \dots, H_M^T T]_{ML \times 1}^T \quad (10)$$

where H_{corr} is the hidden correlation matrix, T_h is hidden layer target matrix and β_{ens} is the global output matrix.

$$H_{corr} = \begin{bmatrix} C_1 H_1^T H_1 + 2\alpha_1 & \dots & C_2 H_1^T H_M \\ C_2 H_2^T H_1 & \dots & C_2 H_2^T H_M \\ \vdots & \ddots & \vdots \\ C_2 H_M^T H_1 & \dots & C_1 H_M^T H_M + 2\alpha_M \end{bmatrix} \quad (11)$$

The global output matrix can be derived from (9):

$$\beta_{ens} = H_{corr}^{-1} T_h \quad (12)$$

Online Ensemble Random Weights Neural Network: In order to make the ensemble neural network learn the sequential data stream one-by-one or chunk-by-chunk, the online learning method (Ding et al. 2016) is adopted in this paper.

For the existing sample $\{x(k), y(k)\}$ and new sample $\{x(k+1), y(k+1)\}$, according to (11), the new hidden correlation matrix can be denoted as:

$$H_{corr} = \begin{bmatrix} C_1 (H_1^k)^T H_1^k + 2\alpha_1 & \dots & C_2 (H_1^k)^T H_M^k \\ C_2 (H_2^k)^T H_1^k & \dots & C_2 (H_2^k)^T H_M^k \\ \vdots & \ddots & \vdots \\ C_2 (H_M^k)^T H_1^k & \dots & C_1 (H_M^k)^T H_M^k + 2\alpha_M \end{bmatrix} + \begin{bmatrix} C_1 (H_1^{k+1})^T H_1^{k+1} + 2\alpha_1 & \dots & C_2 (H_1^{k+1})^T H_M^{k+1} \\ C_2 (H_2^{k+1})^T H_1^{k+1} & \dots & C_2 (H_2^{k+1})^T H_M^{k+1} \\ \vdots & \ddots & \vdots \\ C_2 (H_M^{k+1})^T H_1^{k+1} & \dots & C_1 (H_M^{k+1})^T H_M^{k+1} + 2\alpha_M \end{bmatrix} \quad (13)$$

That is:

$$H_{corr} = H_{corr}^k + H_{corr}^{k+1} \quad (14)$$

In the similar way, we can derive:

$$T_h = T_h^k + T_h^{k+1} \quad (15)$$

Thus, the model can be updated by hidden layer correlation matrix and target matrix according to (14) and (15), and the output weights are obtained by (12). Fig.5 is the structure chart of online learning ERNN. The algorithm of online learning ERNN is outlined as follows.

Algorithm of online learning ERNN

Inputs: data sets (carbon residual and asphaltene)

Parameters: ensemble size M , number of hidden node L , penalty coefficient λ , regularization parameter α_m .

- 1: Initialize the input weights and hidden node basis;
- 2: Calculate the parameters C_1 and C_2 ;
- 3: Calculate the hidden output matrix of each basic network;
- 4: Calculate the hidden correlation matrix, hidden target matrix;
- 5: Get the global output matrix by (12);
- 6: Return the ensemble model and calculate outputs and error;
- 7: If new data, go to step 3; else, go to step 10
- 8: Update the hidden correlation matrix, hidden target matrix;
- 9: Go back to step 5;
- 10: End.

4. EXPERIMENTAL RESULT

4.1 Experimental Setup

In this paper, 595 groups of NMR spectra and the corresponding carbon residual and asphaltene data were collected from the refinery in south China between May 2016 and December 2016, of which 495 groups collected from May 2016 to November 2016 were used as the training data, and the remaining 100 groups collected in December 2016 were used for the model validation.

The sigmoid function is chosen as the activation function in ERNN:

$$G(x) = \frac{1}{1 + e^{-x}} \quad (16)$$

The optimal parameters of ERNN are determined by exhaustive linear search. The number of hidden nodes L was searched between 100 and 300 with step 10, the ensemble size M was searched between 3 and 10 with step 1. Similarly, the penalty term coefficient λ and the regularization coefficient α_m were searched in range $[0, 0.5]$ with step 0.01 and in range $[0, 1]$ with step 0.05, respectively. Finally, the optimal parameters are $L^*=200$, $M^*=6$, $\lambda^*=0.09$ and $\alpha_m^*=0.55$. In order to show the effectiveness of the proposed ERNN, experiments are carried out to compare ERNN with DNNE (Alhamdoosh and Wang, 2014), LS-SVM, PLS and MLR.

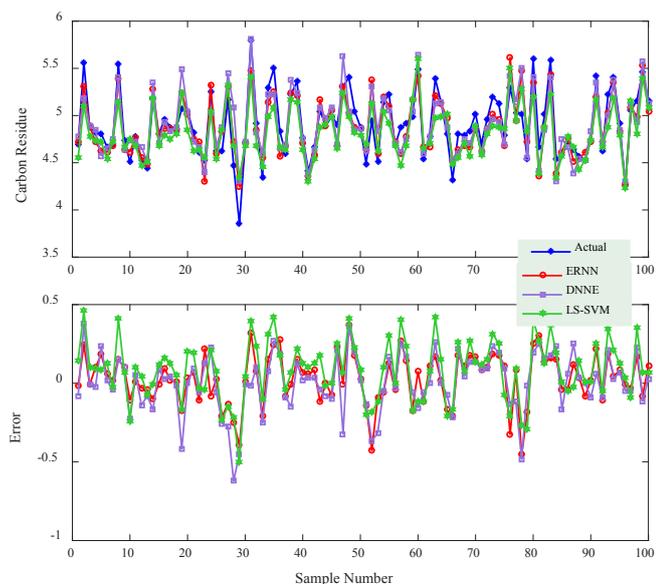


Fig.6. Prediction results and errors of carbon residual of ERNN, DNNE and LS-SVM.

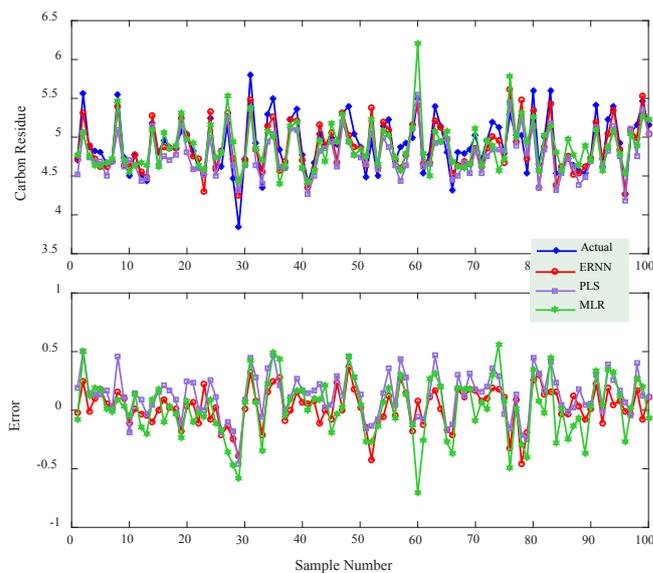


Fig.7. Prediction results and errors of carbon residual of ERNN, PLS and MLR.

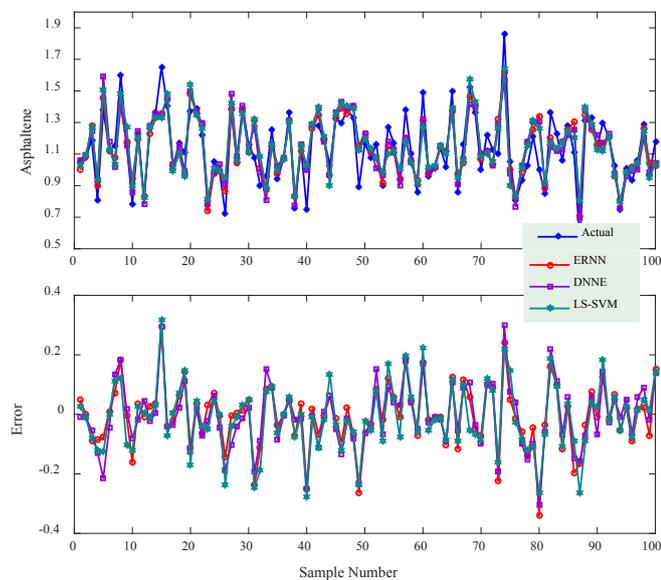


Fig.8. Prediction results and errors of asphaltene of ERNN, DNNE and LS-SVM.

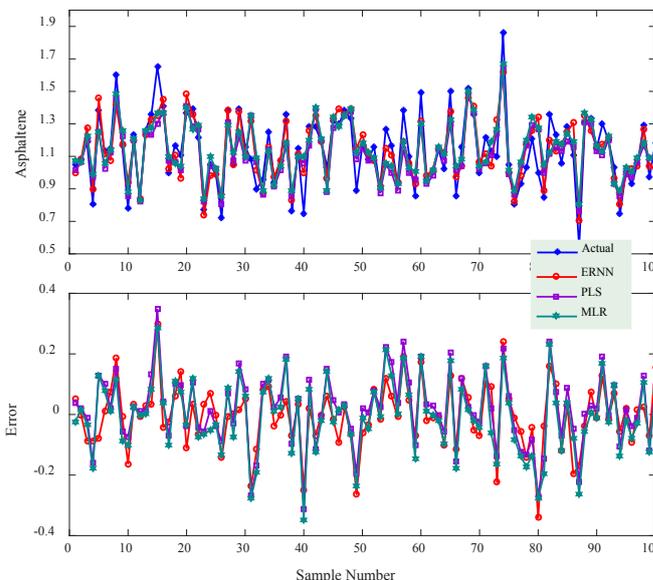


Fig.9. Prediction results and errors of asphaltene of ERNN, PLS and MLR.

Table 1. Efficiency comparison of ERNN, LS-SVM, DNNE, PLS and MLR

Method	Testing time(s)	
	Average	STD
PLS	0.1766	0.0062
MLR	1.9861	0.0241
LS-SVM	0.0915	0.0027
DNNE	0.4153	0.0076
ERNN	0.5733	0.0053

4.2 Comparison and Discussion

All experiments are conducted ten times. Results of the ERNN based model and the comparative models based on DNNE, LS-SVM, PLS and MLR of carbon residual and

asphaltene are shown in Figs.6-10. Fig.6 is the prediction results and prediction errors of carbon residual of ERNN, DNNE and LS-SVM. Fig.7 is the prediction results and prediction errors of carbon residual of ERNN, PLS and MLR. From Figs.6 and 7, it can be seen that the prediction errors of carbon residual of ERNN are almost smaller than that of DNNE, LS-SVM, PLS and MLR. Thus, comparing with the data-based nonlinear models (DNNE, LS-SVM) and the data-based linear model (PLS, MLR), ERNN fits the data better and has the higher prediction accuracy. Figs.8 and 9 show the results of asphaltene of ERNN and the four comparative algorithms. From Figs.8 and 9, it can be observed that except the beginning and the end of the testing data, the prediction errors of ERNN are smaller than that of other involved data-based models on the whole. In addition, we report the testing

Table 2. Performance comparison of ERNN, LS-SVM, DNNE PLS and MLR

Modelling objective	Method	Result	
		RMSE	MAPE
Carbon residual	PLS	0.2461±-	3.903±-
	MLR	0.2272±-	3.749±-
	LS-SVM	0.2053±0.0023	3.364±0.020
	DNNE	0.1833±0.0025	2.868±0.028
	ERNN	0.1613±0.0018	2.604±0.022
Asphaltene	PLS	0.1213±-	9.031±-
	MLR	0.1195±-	8.526±-
	LS-SVM	0.1129±0.0006	8.223±0.009
	DNNE	0.1095±0.0012	7.727±0.014
	ERNN	0.0986±0.0010	7.025±0.011

time of ERNN, DNNE, PLS and MLR in the Table 1, and also calculate the averages and the standard deviations (STD) of root mean square error (RMSE) and mean absolute percentage error (MAPE) of ten experiments to qualify the performance of the involved models in Table 2. The missing values (marked as ‘-’) in the table are the algorithms which the standard deviations are zeroes. Table 1 indicates that the testing time of ERNN is smaller than that of MLR, but it is a little larger than that of DNNE, PLS and LS-SVM. However, the testing time of ERNN is still acceptable. From table 2, we can find that both the RMSEs and MAPEs of the carbon residual and asphaltene of ERNN are smaller than that of DNNE, LS-SVM, PLS and MLR. From the comparisons above, it can be observed that the proposed ERNN outperforms the other algorithms involved in this paper due to the nonlinear approximation ability of ERNN and the introduced regularization term in the ensemble.

5. CONCLUSIONS

Fast evaluation of physical properties of crude oil is an important part in the oil refining industry, while the effective prediction model of physical properties of crude oil is the key to the fast evaluation of crude oil. Carbon residual and asphaltene are considered as two significant indices in crude oil blending, production planning and scheduling and process control of oil refinery. The prediction model of carbon residual and asphaltene based on regularized negative correlation ensemble random weights neural network is presented in this paper, and we use online learning to update the hidden correlation matrix and hidden target matrix by using new data. Furthermore, experiments are carried out using the data collected from a refinery and compared with DNNE, LS-SVM, PLS and MLR. The results demonstrate the effectiveness of the proposed model and its significance of practical application.

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China under Grand 61590922, 61525302, the Project of Ministry of Industry and Information Technology of China under Grand 20171122-6, the Projects of Shenyang under Grand Y17-0-004, and the Fundamental Research Funds for the Central Universities under Grand N160801001, N161608001.

REFERENCES

- Baruník, J. and Malinská, B. (2015). Forecasting the term structure of crude oil futures prices with neural networks. *Applied Energy*, 164, 366-379.
- Sriram Satya., R. M. R., M. D. D. and Hanson, F. V. (2007). Estimation of properties of crude oil residual fractions using chemometrics. *Energy & Fuels*, 21(2), 998-1005.
- He, K., Qian, F., Cheng, H. and Du, W. (2015). A novel adaptive algorithm with near-infrared spectroscopy and its application in online gasoline blending processes. *Chemometrics & Intelligent Laboratory Systems*, 140, 117-125.
- Lin, L. M., Chen, J. and Jin, J. J. (2004). Nuclear magnetic resonance on-line analysis technology and its application in petrochemical unit. *Automation In Petro-chemical Industry*.
- Behera, B., Ray, S. S. and Singh, I. D. (2008). Structural characterization of fcc feeds from indian refineries by nmr spectroscopy. *Fuel*, 87(10-11), 2322-2333.
- Barbosa, L. L., Sad, C. M. S., Morgan, V. G., Figueiras, P. R. and Castro, E. R. V. (2016). Application of low field nmr as an alternative technique to quantification of total acid number and sulphur content in petroleum from brazilian reservoirs. *Fuel*, 176, 146-152.
- Shea, T. and SeldaGunsel. (2003). Modeling base oil properties using NMR spectroscopy and neural networks. *A S L E Transactions*, 46(3), 296-302.
- Igel'nik, B. and Pao, Y. H. (1995). Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Transactions on Neural Networks*, 6(6), 1320-9.
- Hansen, L. K. and Salamon, P. (1990). Neural network ensembles. *Pattern Analysis & Machine Intelligence IEEE Transactions on*, 12(10), 993-1001.
- Chen, H. and Yao, X. (2009). Regularized negative correlation learning for neural network ensembles. *IEEE Transactions on Neural Networks*, 20(12), 1962-79.
- Brown, G., Wyatt, J., Harris, R. and Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Information Fusion*, 6(1), 5-20.
- Chong, S. Y., Tino, P. and Yao, X. (2009). Relationship between generalization and diversity in coevolutionary learning. *IEEE Transactions on Computational Intelligence & Ai in Games*, 1(3), 214-232.
- Wang, S. and Yao, X. (2012). Relationships between diversity of classification ensembles and single-class performance measures. *IEEE Transactions on Knowledge & Data Engineering*, 25(1), 206-219.
- Liu, Y. and Yao, X. (1999). Ensemble learning via negative correlation. *Neural Networks the Official Journal of the International Neural Network Society*, 12(10), 1399-1404.
- Ding, J., Wang, H., Li, C., Chai, T. and Wang, J. (2016). An online learning neural network ensemble with random weights for regression of sequential data stream. *Soft Computing*, 1-19.
- Alhamdoosh, M. and Wang, D. (2014). Fast decorrelated neural network ensembles with random weights. *Information Sciences*, 264(6), 104-117.