Energy 216 (2021) 119299

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Energy futures price prediction and evaluation model with deep bidirectional gated recurrent unit neural network and RIF-based algorithm

Bin Wang*, Jun Wang

Institute of Financial Mathematics and Financial Engineering, School of Science, Beijing Jiaotong University, Beijing, 100044, PR China

ARTICLE INFO

Article history:
Received 10 February 2020
Received in revised form
7 October 2020
Accepted 7 November 2020
Available online 17 November 2020

Keywords:
Energy market futures prices
Prediction model
Random inheritance formula
Gated recurrent unit
Deep bidirectional learning
Synchronization evaluation

ABSTRACT

Energy resources have firmly occupied an unshakable position, which is indispensable both in industrial field and daily life. More accurate prediction of energy futures price has always been a challenging issue. Motivated by this problem, a novel random deep bidirectional gated recurrent unit neural network is constructed to achieve more accurate forecasts of international crude oil futures prices. The random inheritance formula is proposed and integrated into the training process of the model, and it reflects the timeliness of historical data. Both the random inheritance formula and the deep bidirectional learning can effectively improve the model's acquisition of effective information from historical data and improve the model's accuracy. The proposed model is compared with SVM, GRU, ERNN, LSTM, DBGRUNN and RIF-GRUNN models, and a variety of evaluation indicators as well as a novel synchronization evaluation method of q-DSCID are used to measure accuracy. The empirical research results of four crude oil futures prices and coarse-grained moving absolute returns show that the proposed model outperforms the comparison models. For the Brent crude oil futures price prediction, its metrics R^2, MAE, TIC, RMSE and SMAPE are 0.998, 0.200, 0.002, 0.267 and 0.283, which are the best in the comparison models.

1. Introduction

Crude oil is an indispensable source of energy in industrial and agricultural production, and it is mainly engrained in fuels, chemical raw materials and other purposes [1]. Crude oil has an irreplaceable impact on the daily life of the people as extensive as the country's economy, politics and military [2]. Fluctuations in crude oil prices can trigger a country's economic recession and even an economic crisis, causing wars and conflicts between nations [3,4]. Therefore, how to achieve a more accurate forecast of crude oil futures prices has become an increasingly focused issue. Kristjanpoller and Minutolo [5] proposed a hybrid model ANN-GARCH to predict oil price return volatility, and the results indicate that the ANN improves forecasting accuracy over the GARCH and ARFIMA model prediction. Cheng et al. [6] introduced a new hybrid vector error correction and nonlinear autoregressive neural network (VEC-NAR) model to predict future crude oil prices and verified the accuracy of the model.

Crude oil futures prices have nonlinearities, uncertainties and volatility [7,8]. Consequently, it is challenging to achieve high-precision prediction [9–11]. E tl al [9]. proposed a forecasting model based on VMD, ICA and ARIMA to predict crude oil price. Jammazi and Aloui [10] construct a hybrid model HTW-MPNN to achieve prominent prediction of crude oil price. Price series prediction is a typical time series forecast problem [12,13]. Support vector machine (SVM) and support vector regression (SVR) are employed earlier in settling time series forecast problems [14,15]. They have favorable theoretical basis but are not suitable for large sample price prediction. Therefore, optimization problems based on these two mediums are also increasing. Guo et al. [16] used the support vector machine to predict the oil price. In order to reduce the empirical component selection of the traditional algorithm, the genetic algorithm optimization parameters are used to construct a new hybrid model. The results show that this optimizes the accuracy of the model. Kao et al. [17] applied SVR to predict the stock price series and proposed the NLICA-SVR model. Experimental

* Corresponding author. Beijing Jiaotong University, Beijing, 100044, China.
E-mail addresses: binwangbjtu@bjtu.edu.cn (B. Wang), wangjun@m.bjtu.edu.cn (J. Wang).

https://doi.org/10.1016/j.energy.2020.119299
0360-5442/

results show that the proposed model has better prediction accuracy, which indicates that ICA has better optimization ability for SVR. The powerful self-learning function of artificial neural network (ANN) provides a solution for the prediction of large sample sequences. The results of using artificial neural network to predict price series are often superior to traditional theoretical methods [18–20]. Ekonomou [21] developed ANN model to predict Greek energy consumption. The produced ANN results for years 2005–2008 are compared with the results produced by a linear regression method, a support vector machine method and with real energy consumption records showing a great accuracy. Keles et al. [22] proposed a power price prediction model based on artificial neural network. The clustering algorithm is used to optimize the screening data. Compared with the seasonal ARIMA model, the ANN-model achieved better results. Artificial neural networks possess exceptional manifestation in price forecasting, but lack the temporal correlation between data. The recurrent neural network (RNN) solves this conundrum favourably, and realizes the association between data at different times through the recurrent layer, and ameliorates the network architecture [23–25]. Berradi and Lazaar [26] applied a recurrent neural network to predict the stock price of Casablanca Stock Exchange, and used principal component analysis to reduce the number of features and improve the prediction accuracy of RNN. Wang and Wang [27] constructed a new predictive model ST-ERNN to forecast the crude oil prices as well as oil stock prices. The predicted values of the proposed model on the crude oil (and oil stock) prices are more in agreement with the real values than other models. In-depth research has found that recurrent neural networks have long-term dependencies in training and vanish gradient problems, leading to a decline in model prediction exactness.

With the advent of the era of big data, machine learning algorithms have been better evolved. Gated recurrent unit (GRU) was proposed as a more sophisticated and reliable machine learning algorithm. The long-term dependence of RNN and the disappearance of gradients are solved magnificently [28]. GRU is a variant of Long short-term memory (LSTM) [29]. Compared with LSTM, GRU has a more compact configuration and higher efficiency. Studies have also shown that the effects of the two in practical applications are often not much different. Thanks to its superior performance, GRU has been widely used in various fields, and its hybrid optimization model has gradually increased. Wu et al. [30] applied GRU network to forecast short-term load considering impact of electricity price. The simulation study shows that this method can effectively improve the accuracy of short-term load forecasting compared with traditional methods. Chen et al. [31] proposed a general two-step method for residual life prediction, first applying KPCA to extract features and then applying GRU prediction. The results show that GRU is better than LSTM in training time and prediction accuracy, and can provide better prediction for nonlinear degradation process of complex systems.

Recently, many articles on energy price series forecasting have been published, providing more methods for energy price prediction. Hooman and Seyed [32] introduced a novel hybrid model employed ANFIS, ARFIMA, and Markov-switching to forecast Brent crude oil price and results show genetic algorithm weighted hybrid model outperforms the other models. Qiao and Yang [33] proposed WT-SAE-LSTM to forecast electricity prices of USA and results show that the proposed model outperforms other AI models, such as back propagation neural network, in forecasting accuracy. Cen and Wang [34] established a new prediction model with long short term memory based on prior knowledge data transfer to predict crude oil price and empirical results show that data transfer can greatly improve prediction accuracy of long short term memory. E et al. [35] proposed a novel hybridization of multi-scale model for predicting the energy price based on independent component analysis, gated recurrent unit neural network and support vector regression and experiments demonstrate the validity and reliability of the improved model. Most of them are based on improvements or mixtures of existing models and methods to achieve the goal of further improving the accuracy of model forecasts. Inspired by this, this paper introduces a novel energy price prediction model that integrates GRU, deep learning and bidirectional learning structure, and proposes a random inheritance formula to optimize the model training process.

In the process of applying neural network model to predict the price series of crude oil futures, it is necessary to provide sufficient historical data for model learning and training. Therefore, how to implement better data effective information extraction and more reasonable information application is a tough challenge. The extraction and utilization of input information directly affects the accuracy of the model prediction results. In order to counter this problem and improve the performance of the model, we construct a novel random deep bidirectional gated recurrent unit neural network (RIF-DBGRUNN) model. First, the deep learning method is applied to increase the depth and ameliorate the predictive property of the model. Then the bidirectional learning structure is established, which consists of two independent GRU units, input information is trained in both directions to obtain more effective information. This achieves a more comprehensive feature extraction of input information and improves the learning effect of the model. Some research found that the application value of the data points of the input sequence is different [36]. Future price fluctuations will inherit the characteristics of some historical data, more or less, depending on the time interval between the predicted sample points and the past sample points [37,38]. Based on this characteristic, a random inheritance formula (RIF) is proposed based on geometric Brownian motion [39,40]. In the process of model learning, the gradient correction is carried out for each learning, and the random inheritance formula is applied to the gradient correction. The gradient correction method based on RIF can correct the parameter matrix more reasonably and optimize the prediction result.

Aapply the RIF-DBGRUNN to predict four sets of energy futures price series, compare the prediction results with six generally utilized models, and perform linear regression analysis, basic accuracy measurement and q-order dyadic scales complexity invariant distance (q-DSCID) evaluation on the prediction results [41,42]. Comparing the model on a GRU basis and calculating the increase percentage (IP) in metrics. The proposed model is also applied to predict coarse-grained moving absolute return (CMAR) series. All test results demonstrate the excellent accuracy and stability of the proposed model.

The main contribution of this article can be summarized as four points. First, a novel RIF-DBGRUNN model is established to achieve more accurate energy futures prices prediction. Second, the random inheritance formula is introduced and integrated into the training process of GRU model to optimize the model training process. Third, deep bidirectional learning is applied to model construction in order to capture more effective historical information. Fourth, q-DSCID, a more reliable synchronization evaluation method, is used in the comparison and evaluation of model accuracy. The remaining part of this paper includes the following sections. Section 2 introduces the adopted method and the integrated framework of the proposed model. Section 3 describes the basic and complex metrics. Section 4 shows the experiment and comparative analysis results for crude oil futures price and coarse-grained moving absolute return series. Section 5 revels conclusions and expectations.

2. Methodology

2.1. Gated recurrent unit

GRU model is an evolution of traditional RNN model. It solves the problem of long-term dependence in traditional RNN network, and has the outstanding characteristics of fast training and high precision [43]. Its structure unit is shown in Fig. 1. GRU can be regarded as a variant of LSTM. Based on LSTM network, GRU improves the activation function of hidden layer nodes and uses the activation function controlled by gate valve to calculate the output of hidden layer, so that each threshold control unit can adaptively capture the information of different time scales and simplify the operation process inside the model. It solves the problems of complicated calculation and low efficiency in LSTM. The difference between them is that LSTM contains three function gates: input gate, forgetting gate and output gate to control input value, memory value and output value, while GRU contains two function gates, which are update gate z_t and r_t reset gate, respectively. The update gate is used to control the extent to which the state information of the previous moment is brought into the current state, and the reset gate controls how much information of the previous state is stored in the current candidate set c_t. Experiments show that the performance of GRU and LSTM is similar, but compared with LSTM, GRU has the advantages of easy convergence, simple structure and fast training. The forward propagation process of GRU is expressed as [44].

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \\ z_t &= \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \\ c_t &= \tanh(W_c \cdot [r_t * h_{t-1}, x_t] + b_h) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * c_t \\ y_t &= \sigma(W_o h_t) \end{aligned} \tag{1}$$

where $[\cdot, \cdot]$ denotes the connection of two vectors and * refers to dot product operation of matrices, "$\cdot$" represents two matrices multiplied by their elements, x_t is the sample point input at time t, c_t is the candidate set, h_t is the state information flowing into the next layer, and y_t is the output result of input. W_r, W_z, W_c represent the parameter matrix from $[h_{t-1}, x(t)]$ to r_t, z_t and c_t, respectively, W_o is the connection matrix from hidden layer to output layer. σ and *tanh* are the activation function of *Sigmoid*$(\cdot)$ and *Tanh*$(\cdot)$, whose calculation formula and derivative expression are as follows

$$\begin{aligned} \sigma(x) = y &= \frac{1}{1 + e^{-x}}, & \tanh(x) = y &= \frac{e^x - e^{-x}}{e^x + e^{-x}}. \\ \sigma'(x) &= y(1 - y) & tanh'(x) &= 1 - y^2 \end{aligned} \tag{2}$$

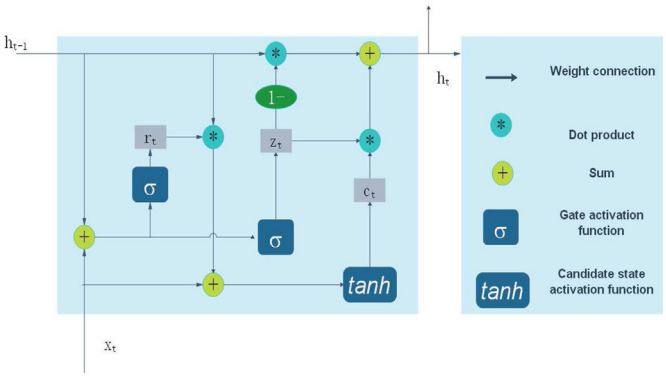

Fig. 1. Architecture of gated recurrent unit.

2.2. Error correction algorithm based on random inheritance formula

Based on geometric Brownian motion, a novel random inheritance formula is proposed to train the neural network model. Research shows that the prediction of market price series needs to integrate a great amount of past data, and historical data has a significant timeliness for future results, that is, recent data has a stronger impact factor than earlier data. Therefore, how to take into account as much as possible the impact factors of historical data on the prediction results, and how to distinguish the impact intensity differences between historical data in different periods reasonably is an important thinking angle to ameliorate the prediction preciseness of the model. Consequently, we propose a random inheritance formula $\psi(t_n)$ based on geometric Brownian motion to solve this problem. Specific definitions are as follows

$$\psi(t_n) = \frac{1}{\gamma} \exp\left\{ \int_{t_0}^{t_n} \nu(t)dt + \int_{t_0}^{t_n} \zeta(t)dB(t) \right\} \tag{3}$$

$$\nu_t = 1 \Big/ \exp\{at\}, \zeta(t) = \left[\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \right]^{\frac{1}{2}} \tag{4}$$

where $\gamma(>0)$ represents the time intensity coefficient, which is equal to 2, t_0 denotes the time when the latest data occurs and t_n is the time of any historical data. N is the total length of the input $x_t(t = 1, 2, \cdots, N)$ and represents the mean of input samples. $\nu(t)$ is a drift function, which is used to simulate the overall trend of events. The parameter a is related to the historical price of crude oil futures, and a large number of preset experiments can be used to select the group with the best prediction result as the value of a. $B(t)$ represents standard Brown motion which indicates that the future trajectory of particles in a given time does not depend on the occurrence of past events. $\zeta(t)$ is a volatility function, together with $B(t)$ represents unmeasurable factors. The stochastic inheritance function can well combine the historical data characteristics of crude oil futures prices, increase the weight in the recent data retraining process, reduce the older data weight, and then improve the prediction accuracy of the model. Next, the RIF is applied to the GRU training process to optimize parameter updating.

Eq. (3) gives the forward propagation process of GRU, in which W_r, W_z and W_c are mosaic and can be differentiated as follows

$$\begin{aligned} W_r &= W_{rx} + W_{rh} \\ W_z &= W_{zx} + W_{zh} \\ W_c &= W_{cx} + W_{ch} \end{aligned} \tag{5}$$

where W_{ij} denotes that W_{ij} is the connection matrix from j to i. The input and output of the reset gate, update gates, candidate sets and output layer can be expressed as

$$net_r^t = x^t W_{rx} + h^{t-1} W_{rh}, r_j^t = \sigma(net_r^t)^j \tag{6}$$

$$net_z^t = x^t W_{zx} + h^{t-1} W^{zh}, z_j^t = \sigma(net_z^t)^j \tag{7}$$

$$net_c^t = x^t W_{cx} + \left(r^t * h^{t-1}\right) W_{ch}, c_j^t = \tanh(\text{net}_c^t)^j \tag{8}$$

$$net_y^t = h^t W_o, y_j^t = \sigma\left(\text{net}_y^t\right)^j. \tag{9}$$

Define t_n as the time of sample point n $(n = 1, 2, \cdots, N)$. The error

of the sample point at time t_n is defined as

$$E_{t_n} = \frac{1}{2}\epsilon_{t_n}^2 = \frac{1}{2}(o_{t_n} - y_{t_n})^2. \tag{10}$$

where y_{t_n} and represent the predicted and true values at time t_n. Combining RIF with error correction process, a new definition of error E_{t_n} can be obtained, which is expressed as follows

$$E_{t_n} = \frac{1}{2}\psi(t_n)\epsilon_{t_n}^2 = \frac{1}{2}\psi(t_n)(o_{t_n} - y_{t_n})^2 \tag{11}$$

where $\psi(t_n)$ denotes the RIF, which will assign corresponding weight on the basis of the time of sample occurrence. Then the global error can be defined as

$$E = \frac{1}{N}\sum_{n=1}^{N} E_{t_n} = \frac{1}{2N}\sum_{n=1}^{N}\frac{1}{\gamma}\exp\left\{\int_{t_0}^{t_n}\nu(t)dt + \int_{t_0}^{t_n}\zeta(t)dB(t)\right\}(s_{t_n} - y_{t_n})^2. \tag{12}$$

Applying the redefined error function E, the parameter matrix is updated by gradient descent method. First, the partial derivatives of the error function with respect to each parameter are obtained, and then the parameters are updated. The specific process is as follows

$$\begin{cases} \dfrac{\partial E}{\partial W_o} = \delta_{y,t}\psi(t)h_t \\ \dfrac{\partial E}{\partial W_{zx}} = \delta_{z,t}\psi(t)x_t, \quad \dfrac{\partial E}{\partial W_{zh}} = \delta_{z,t}\psi(t)h_{t-1} \\ \dfrac{\partial E}{\partial W_{cx}} = \delta_t\psi(t)x_t, \quad \dfrac{\partial E}{\partial W_{ch}} = \delta_t\psi(t)(r_t{*}h_{t-1}) \\ \dfrac{\partial E}{\partial W_{rx}} = \delta_{r,t}\psi(t)x_t, \quad \dfrac{\partial E}{\partial W_{rh}} = \delta_{r,t}\psi(t)h_{t-1} \end{cases} \tag{13}$$

where the parameters are expressed as

$$\begin{cases} \delta_y^t = (o^t - y^t){*}f'\left(net_y^t\right) \\ \delta^t = \delta_h^t{*}z^t{*}g'\left(net_h^t\right) \\ \delta_z^t = \delta_h^t{*}\left(c^t - h^{t-1}\right){*}f'\left(net_z^t\right) \\ \delta_r^t = h^{t-1}{*}\left[\left(\delta_h^t{*}z^t{*}g'\left(net_h^t\right)\right)W_{ch}\right]{*}f'\left(net_r^t\right) \\ \delta_h^t = \delta_y^t W_o + \delta_z^{t+1}W_{zh} + \delta^{t+1}W_{ch}{*}r^{t+1} + \delta_r^{t+1}W^{rh} + \delta_h^{t+1}{*}\left(1 - z^{t+1}\right) \end{cases} \tag{14}$$

where $f(\cdot)$ and $g(\cdot)$ are respectively expressed as activation function of and $tanh(\cdot)$. The updating rule of the parameter matrix W_{ij} is

$$W_{ij} = W_{ij} - \eta\cdot\frac{\partial E}{\partial W_{ij}}. \tag{15}$$

with continuous iteration, the parameter matrix is updated continuously, so that the output results are closer to the real data.

2.3. Deep bidirectional GRU neural network

Based on the superiority of the GRU introduced in Section 2.1, we choose it as the unit of the recurrent neural network to build a GRUNN model by replacing the hidden layer in the RNN with the GRU unit. Its structure is shown in Fig. 2. Replacing the hidden layer in traditional RNN with GRU to optimize the network structure more precisely. The training process can be expressed as

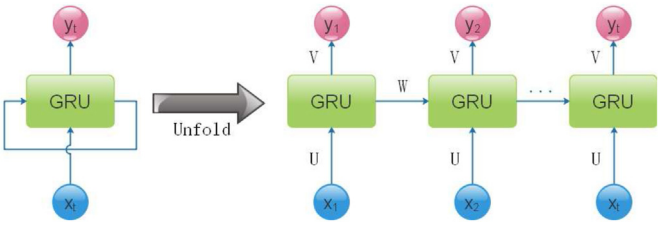

Fig. 2. Topological architecture of the GRU with RNN.

$$y_t = Vf(Ux_t + Wf(Ux_{t-1} + Wf(Ux_{t-2} + Wf(Ux_{t-3} + \cdots)))) \tag{16}$$

where U, V and W represent the parameter matrix between input layer, GRU, output layer and h, respectively.

The research shows that increasing the depth of the neural network model can effectively improve the training accuracy of the model [45]. The core idea of a deep recurrent neural network is to make the input pass through more non-linear layers to improve the network's expression and learning capabilities. Fig. 3(a) shows a three-layer deep GRUNN model structure which is similar to GRUNN, except that the input of each time step of the deep neural network will be processed by three GRU units. In the deep GRUNN architecture as shown in Fig. 3(a), the input at time t, x_t is introduced to the first unit GRU_1 along with the previous hidden state, the output of GRU_1 is calculated as shown in section 2.1 and goes forward to the next time step and also goes upward to the unit GRU_2. The GRU_2 uses the hidden state of GRU_1 along with the previous hidden state to compute the output, that goes forward to the next time step and upward to the unit GRU_3. The output of each GRU training layer has a temporal correlation, that is, the output of each GRU unit flows into the next time step GRU unit while flowing into the next layer of GRU unit.

Crude oil futures prices are continuous and closely related to historical data. The prediction results depend on the entire training set, so we consider the bidirectional training model and construct the bidirectional GRUNN as shown in Fig. 3(b). It combines two GRUNN network layers: one GRUNN moving from the beginning of the sequence and the other moving from the end of the sequence. The basic idea is: for each training sequence, two GRUNN models are established in the forward and backward directions, and the hidden layer nodes of the two models are connected to the same output layer. The special data processing method can provide complete historical and future information for each time point in the input sequence of the output layer. This method not only improves the effective information acquisition of the model, but also makes it more flexible and reliable.

To further improve the accuracy of the model, we integrated the deep GRUNN model and the bidirectional GRUNN model to build a deep bidirectional gated recurrent unit neural network (DBGRUNN) model. It combines the advantages of the deep GRUNN model and the bidirectional GRUNN model, which improves the model's learning and expression capabilities while fully interpreting the original data. Fig. 3(c) shows the architecture of the three-layer DBGRUNN model. The first two GRU units are information correction extraction layer, and both outputs are the input of the second layer of unit GRU_3. The output after training through the second layer GRU_3 is passed to the third layer of unit GRU_4 and the next

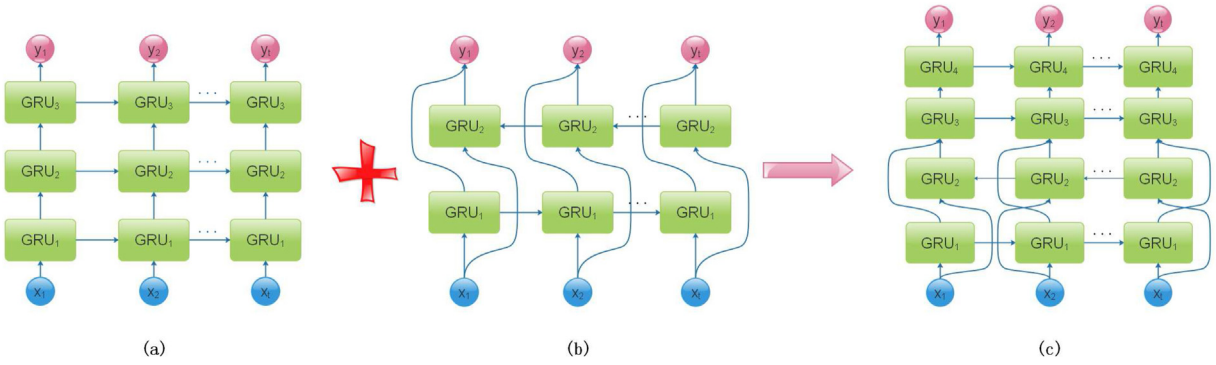

Fig. 3. (a) Architecture of deep GRUNN. (b) Architecture of bidirectional GRUNN. (c) Architecture of deep bidirectional GRUNN.

time step.

2.4. Integrated framework of the proposed model

The error correction method based on RIF is applied to the training process of DBGRUNN, and a novel RIF-DBGRUNN model is constructed, which is shown in Fig. 4. The specific process can be described as follows. (a) Selecting sample set, screening and processing sample set. (b) Build the deep GRUNN and bidirectional GRUNN models, then obtain the optimized deep bidirectional GRUNN. (c) Constructing the novel RIF-DBGRUNN by combining RIF-based error correction algorithm. (d) Training the model to obtian the prediction results.

After selecting the data set, we normalize the data and use the normalized data as the input of the model. Construct the RIF-DBGRUNN model, and realize the building and training of the model through Tensorflow. The number of moving window is selected as 5, which means that the moving window will use the data of the previous four days to predict the data of the fifth day. Set initial parameters and build a neural network model. In the training process of the model, we choose the root mean square error (RMSE) to measure the results of the training set and the test set, which can reflect the consistency between the prediction results of the model and the real data after each training under the current parameter settings. When the training is over, de-normalize the results to draw a comparison figure between the final output results and the real data. Next, change the selection of hyperparameters to optimize settings by combining RMSE indicators and visualization of prediction results. After a large number of experiments, we finally choose the RIF-DBGRUNN model with a depth of three layers, respectively the bidirectional training layer and two GRU unit layers, in which the hidden size of each GRU unit is 30, and the batch size is 33. To assess the predicted results of RIF-DBGRUNN, they are compared with those of SVM, GRU, ERNN, LSTM, DBGRUNN and RIF-GRUNN models.

3. Performance evaluation

3.1. Basic accuracy metrics

To further study the predictive abilities of RIF-DBGRUNN, some regression model evaluation indicators are utilized in this work. Let $o_t$denotes the true value, $y_t$is the predicted value and $\bar{o}$represents the mean of samples. The selection of indicators and their formulas

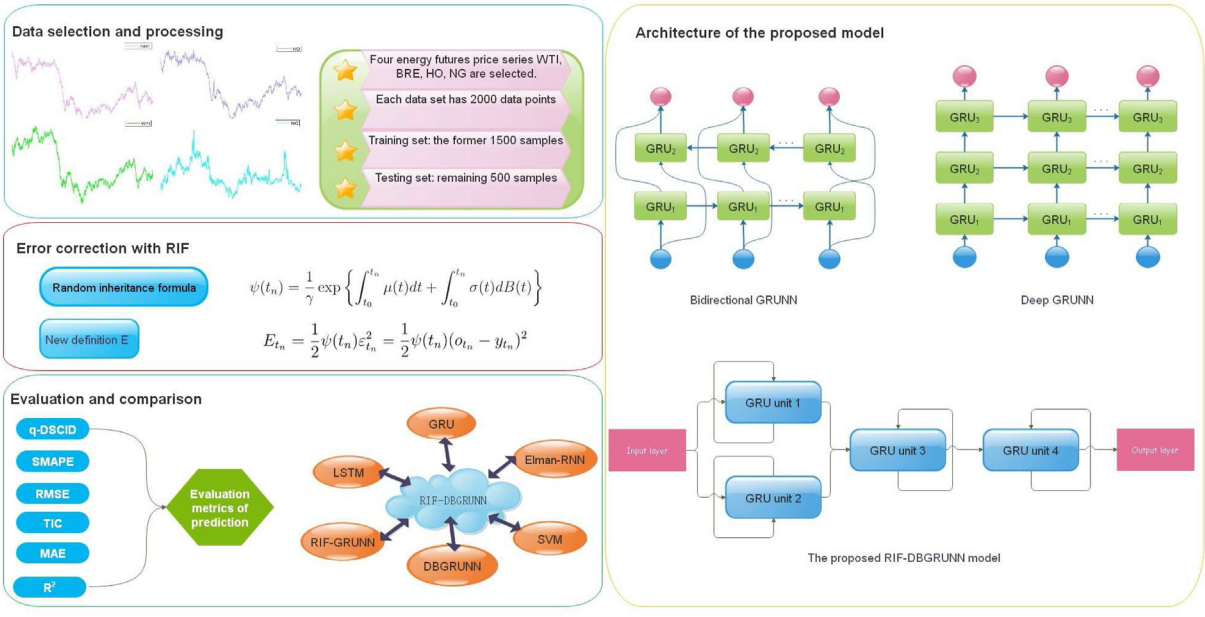

Fig. 4. Integrated architecture framework of the RIF-DBGRUNN model.

Table 1
Selection and introduction of multiple evaluation indicators.

Comprehensive evaluation indicators	
Mean absolute error (MAE)	$\frac{1}{N}\sum_{t=1}^{N}\lvert o_t - y_t\rvert$
Root mean square error (RMSE)	$\sqrt{\frac{1}{N}\sum_{t=1}^{N}(o_t - y_t)^2}$
Symmetric mean absolute percentage error (SMAPE)	$100\times\frac{2}{N}\sum_{t=1}^{N}\frac{\lvert o_t - y_t\rvert}{\lvert o_t\rvert + \lvert y_t\rvert}$
Theil inequality coefficient (TIC)	$\frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(o_t - y_t)^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}o_t^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N}y_t^2}}$
R-squared (R^2)	$1 - \frac{\sum_{t=1}^{N}(o_t - y_t)^2}{\sum_{t=1}^{N}(y_t - \bar{o})^2}$

are listed in Table 1 [46–49]. Through these objective evaluation indicators, the comprehensive performance of the model is evaluated and the predictive ability of each model is compared.

3.2. Complex accuracy metrics

In this section, a competent evaluation method q-order dyadic scales complexity invariant distance (q-DSCID) is presented for energy futures price time series, which can test and quantify the correlation between different time series. Therefore, we can apply it to weigh the predictions. Suppose two time series $O = \{o_1, o_2\cdots, o_n\}$, $Y = \{y_1, y_2\cdots, y_n\}$. The distance between two sets of sequences is determined as follows

$$\mathrm{ED}^q(O,Y) = \Big[\sum_{i=1}^{n}(o_i - y_i)^q\Big]^{\frac{1}{q}}. \tag{17}$$

The definition of invariant distance can be obtained from the following correlation factors and the distance $\mathrm{EDED}^q(O,Y)$ between two sets of time series O and Y

$$q-\mathrm{CID}(O,Y) = \mathrm{ED}^q(O,Y)\times \mathrm{CF}^q(O,Y) \tag{18}$$

$\mathrm{CF}^q(O,Y)$ is a complexity correlation factor, which is expressed as

$$\mathrm{CF}^q(O,Y) = \frac{\max\{\mathrm{CE}^q(O), \mathrm{CE}^q(Y)\}}{\min\{\mathrm{CE}^q(O), \mathrm{CE}^q(Y)\}}. \tag{19}$$

Take the time series $Y = \{y_1, y_2\cdots, y_n\}$ as an example, its complexity estimation $\mathrm{CE}^q(Y)$ is defined as

$$\mathrm{CE}^q(T) = \Big[\sum_{i=1}^{n-1}(y_{i+1} - y_i)^q\Big]^{\frac{1}{q}}. \tag{20}$$

The q-DSCID evaluation method combines multiple time scales and multiple displacement scale while evaluating the predictions, which can be expressed as

$$q-\mathrm{DSCID} = (O,Y,\tau,S) = \frac{1}{S}\sum_{s=1}^{S} q-\mathrm{CID}(O(\tau,s), Y(\tau,s)) \tag{21}$$

where, take the double-scale time series $Y(\tau,s)$ as an example, specifically indicated as

$$Y(\tau,s) = \Big\{y_1(\tau,s), y_2(\tau,s), \cdots, y_j(\tau,s)\Big\} \tag{22}$$

where the element $y_j(\tau,s)(j = 1,2,\cdots,(n-S)/\tau)$is represented as

$$y_j(\tau,s) = \frac{1}{\tau}\sum_{i=(j-1)\tau+s}^{j\tau+s-1} y_i, \quad j = 1,2,\cdots,(n-S)\Big/\tau. \tag{23}$$

Different from the previous regression evaluation indexes, the q-CID distance has a close relationship with the correlation complexity of time series O and Y. The q-CID will degenerate into the traditional CID distance when $q = 2$, and the $\mathrm{ED}^q(O,Y)$ distance becomes the universal Euclidean distance. The q-CID distance can be scaled by determining the value of q to distinguish the difference between time series more clearly, especially when the difference between time series is subtle. Without considering the displacement scale S, the initial sequence is obtained especially when $\tau = 1$. After combining the displacement scale S, the q-DSCID can be regarded as the average value of q-$\mathrm{CID}(O(\tau,s), Y(\tau,s))$ under different displacement scales. q-DSCID considers time scale parameter τ and displacement scale parameter S, so we can use q-DSCID to test the discrepancy between O and Y in different time scales and displacement scales. Thus, we can see that q-DSCID is a more comprehensive and reliable error evaluation method.

4. Numerical experimentation on empirical analysis and comparison

4.1. Crude oil futures price selection and processing

A total of four data sets were selected: the crude oil futures data of West Texas Intermediate crude oil (WTI), ICE Brent crude oil (BRE), Heating Oil (HO), Henry Hub Natural Gas (NG). Many investors regard the WTI spot contracts as the benchmark price for measuring crude oil price changes in the international energy market. Brent crude oil futures is the most widely used contract in global crude oil pricing. At present, about seventy percent of spot crude oil trading pricing refers to Brent scale. Some Asian crude oil producers have recently shifted from their regional scale to Brent, and there are two-thirds of the world's oil market crude oil prices related to Brent crude oil, including China's refined oil price reference mechanism. As one of the important products of crude oil, natural gas is applied to all walks of life by virtue of its low cost and high efficiency. Fuel oil is also one of the processed products of crude oil. It has many properties and is widely used. Both are processed products of crude oil and are closely related to crude oil. Therefore, it is also meaningful to study their futures prices. The selection and division of the data set is shown in Table 2. The data selection covers the period from 2012 to 2019. Since June 2014, the international crude oil market has begun to plummet, and by the end of December it has fallen by more than 50%. Correspondingly, a significant decrease can be seen in the figures of comparison between the prediction results and the real data. This decline is

Table 2
Data set selection date and division.

	Data set			
	WTI	BRE	HO	NG
Training	1500	1500	1500	1500
Test	500	500	500	500
Start date	2012.01.19	2011.12.19	2012.01.19	2011.10.19
Expiry date	2019.09.14	2019.09.18	2019.09.05	2019.09.01

mainly due to the influence of supply and demand and various non-market factors such as geopolitics and dollar liquidity. The poor performance of the world's major economies and emerging economies in 2014 leads to a decrease in demand for crude oil. At the same time, the escalation of geopolitical events further causes a sharp drop in oil prices. In addition, the trend of the dollar and some speculation behaviors also contribute to the decline of oil prices.

To eliminate the adverse effects caused by singular sample data, we should filter and normalize the data set appropriately. The specific steps are as follows. (i) The 3-σ principle is applied to eliminate the data whose mean fluctuates above or below three times the standard deviation. (ii) We normalize the data with the method of physical deviation standardization whose conversion function is as follows

$$\tilde{O}(t) = \frac{O(t) - \min O(t)}{\max O(t) - \min O(t)} \tag{24}$$

where $O(t)$ stands for the price of time t, $\min O(t)$ and $\max O(t)$ represent the lowest price and the highest price of training set separately. After we get the output value, the real data can be retrieved through

$$O(t) = \tilde{O}(t)(\max O(t) - \min O(t)) + \min O(t). \tag{25}$$

4.2. Model validation of crude oil futures prices

In this section, four sets of energy futures price series are forecasted by the RIF-DBGRUNN model. The forecasting results are shown in Fig. 5. Simultaneously, the absolute correlation error diagram of the prediction results is also shown in Fig. 5, which can be expressed as $RE(t) = |o_t - y_t|/o_t$. *RE*reflects the degree of deviation between the predicted result and the true value. It is believed that the higher the degree of coincidence between the prediction result image and the real data image, the closer the model prediction result is to the real data. It can be seen from the figure that the fluctuation of the prediction result is basically consistent with the actual data, and the image of the prediction result basically covers the real data, indicating that the prediction accuracy of the model is high. The smaller the value of *RE*, the higher the prediction accuracy of the model. Results show that the values of *RE*in the prediction results of WTI and BRE are all stable within the range of $(0, 0.01)$, and the values are small. Only a few of the data points for the HO, and NG predictions slightly exceeded 0.01. This proves that the proposed model performs very well.

4.3. Comparisons and analysis

In this part, we compare the predictive power of the RIF-DBGRUNN model horizontally and vertically. Six models of SVM, GRU, ERNN, LSTM, DBGRUNN, RIF-GRUNN are selected as comparison models [34,50]. The models are constructed and trained separately, and the prediction results of the four groups of crude oil futures price series are predicted. Fig. 6 shows a comparison of the predicted results of the models with the real data. Through the partially enlarged figure, it can be found that the prediction results of SVM are quite different from the real data, indicating that the prediction results of the model are not good, while the prediction results of GRU, ERNN, and LSTM are generally better than SVM, and the difference between the three is not obvious. For the DBGRUNN, RIF-GRUNN and RIF-DBGRUNN models, their prediction results are highly coincidence with the real data, especially the RIF-DBGRUNN model, which shows that the prediction results of the three models are relatively good. In order to further accurately evaluate the

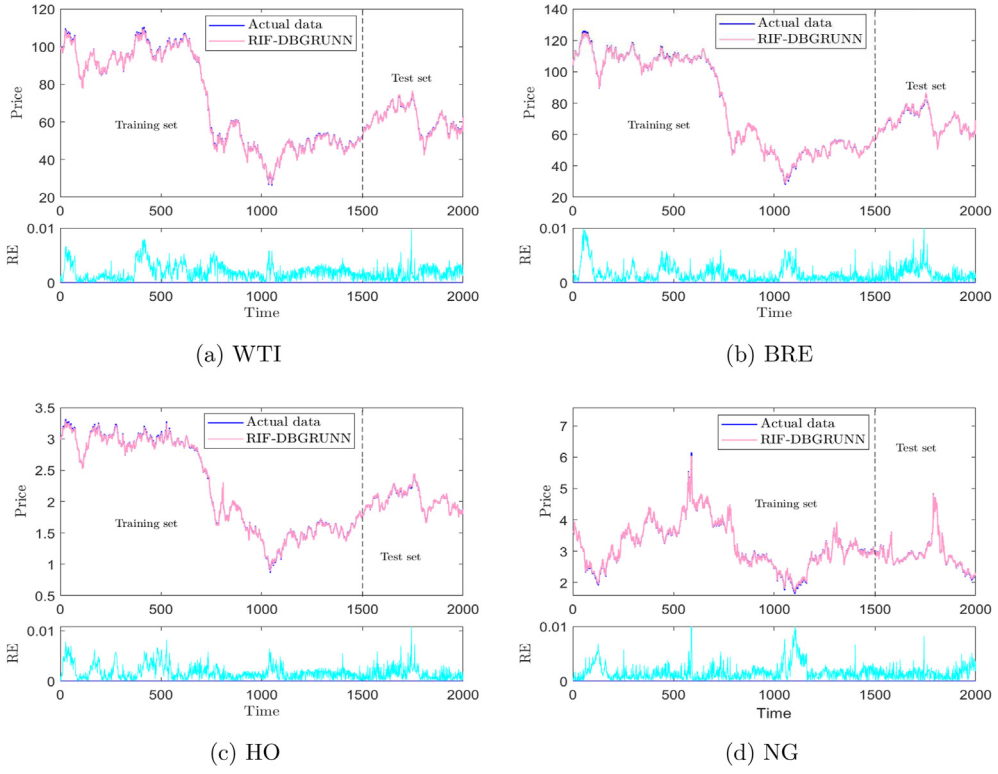

Fig. 5. Prediction results of four sets of futures price series by RIF-DBGRUNN model.

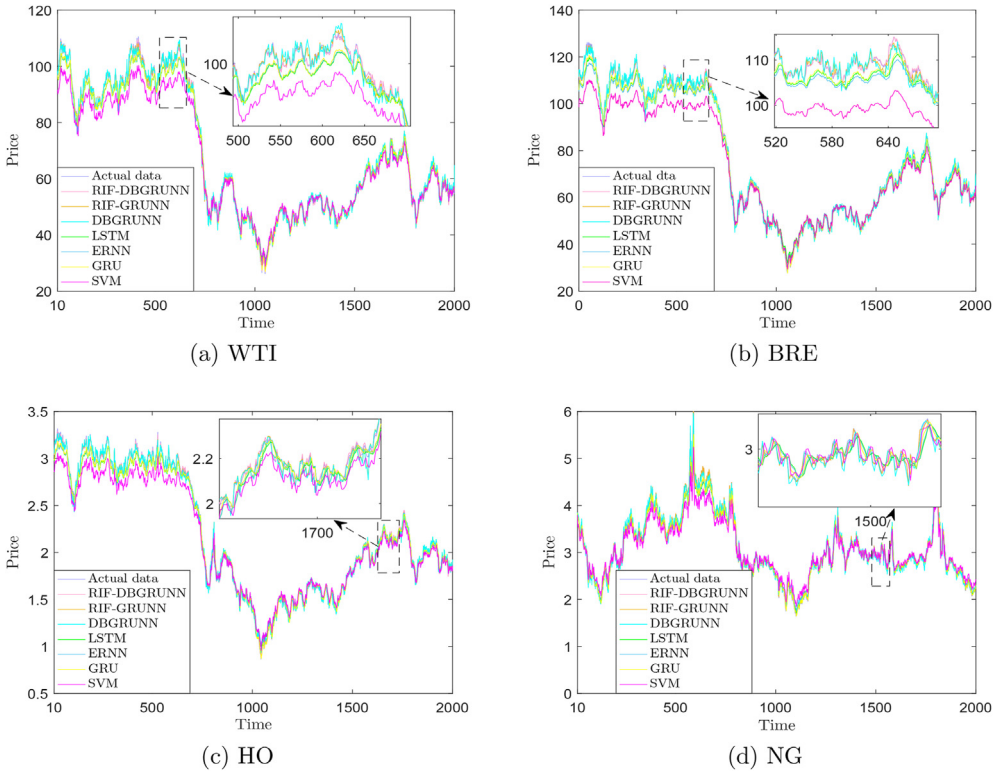

Fig. 6. Multi-model prediction results and real data comparison chart.

prediction accuracy of the model, linear regression analysis and error evaluation of the prediction results of each model are carried out.

Then, a linear regression analysis is performed on the predicted results of the models. Table 3 lists the results of a model fit on the regression equation … The parameters of RIF- DBGRUNN model are very excellent. Take HO as an example, the fitting parameters a and b are 1.025, − 0.048respectively, which is very close to the ideal case of $a = 1$and $b = 0$. However, the fitting parameters of the predicted results of the SVM model are worse than those of other models.

Table 4 gives the basic accuracy metrics of each model for WTI predictions. The value of TIC ranges from 0 to 1, and R^2 reflects the proportion of regression square sum to total square sum. The smaller the TIC value is, the better the prediction result of the model is. The closer the value of R^2 is to 1, the higher the accuracy of the model is. For the convenience of description, we will refer to RIF-DBGRUNN, RIF-GRUNN, DBGRUNN, LSTM, ERNN, GRU, SVM, as Type I, Type II, Type III, Type IV, Type V, Type VI, Type VII. The R^2 value of Type I is the largest equal to 0.999, and the values MAE, TIC, RMSE, SMAPE is minimal, 0.175, 0.002, 0.214, 0.298 respectively.

Table 4
Evaluation indicators of distinct forecasting models for WTI.

Metrics	Model						
	Type VII	Type VI	Type V	Type IV	Type III	Type II	Type I
R^2	0.947	0.960	0.945	0.954	0.983	0.990	0.999
MAE	1.244	1.020	1.213	1.033	0.712	0.551	0.175
TIC	0.013	0.011	0.013	0.012	0.007	0.006	0.002
RMSE	1.574	1.365	1.587	1.400	0.876	0.682	0.214
SMAPE	2.012	1.702	2.022	1.754	1.157	0.894	0.298

This indicates that the prediction result of Type I to WTI is the most accurate in the comparison models. Similarly, Tables 5–7 show the basic accuracy metrics of each model on BRE, HO, and NG. It can be seen that the indicators of Type I are better than other models, indicating that Type I has better learning and prediction capabilities among comparison models.

Further, the complex accuracy metric q-DSCID is applied to evaluate the performance of the proposed model. Choose $S = \text{mean}\{\tau\}$, τ varies from $1, 2, 3, \cdots, 40$, and Type I to Type VII are predicted 50 times respectively. The q-DSCID values of each model

Table 3
Linear regression parameters for different models.

Model	WTI		BRE		HO		NG	
	A	b	a	b	a	b	A	b
SVM	0.873	6.985	0.867	8.030	0.861	0.256	0.768	0.679
GRU	0.969	2.046	0.966	2.540	0.951	0.105	0.915	0.240
ERNN	0.970	1.905	0.979	2.077	0.966	0.071	0.876	0.356
LSTM	0.946	3.191	0.944	3.473	0.937	0.127	0.876	0.366
DBGRUNN	1.001	0.472	1.033	−1.582	0.961	0.088	0.989	0.011
RIF-GRUNN	1.002	0.198	1.024	−1.177	0.969	0.075	0.989	0.032
RIF-DBGRUNN	1.022	−1.428	1.025	−1.555	1.025	−0.048	0.989	0.032

Table 5
Evaluation indicators of distinct forecasting models for BRE.

Metrics	Model						
	Type VII	Type VI	Type V	Type IV	Type III	Type II	Type I
R^2	0.932	0.952	0.929	0.945	0.976	0.986	0.999
MAE	1.462	1.089	1.346	1.105	0.878	0.666	0.200
TIC	0.013	0.011	0.013	0.012	0.008	0.006	0.002
RMSE	1.807	1.512	1.836	1.531	1.063	0.817	0.267
SMAPE	2.123	1.625	1.996	1.691	1.261	0.963	0.283

Table 6
Evaluation indicators of distinct forecasting models for HO.

Metrics	Model						
	Type VII	Type VI	Type V	Type IV	Type III	Type II	Type I
R^2	0.929	0.943	0.928	0.946	0.975	0.978	0.998
MAE	0.034	0.028	0.032	0.027	0.020	0.019	0.005
TIC	0.010	0.009	0.010	0.009	0.006	0.006	0.002
RMSE	0.042	0.037	0.041	0.036	0.024	0.023	0.006
SMAPE	1.677	1.391	1.576	1.395	0.978	0.952	0.227

Table 7
Evaluation indicators of distinct forecasting models for NG.

Metrics	Model						
	Type VII	Type VI	Type V	Type IV	Type III	Type II	Type I
R^2	0.915	0.949	0.933	0.946	0.992	0.999	0.999
MAE	0.087	0.063	0.072	0.062	0.030	0.008	0.008
TIC	0.023	0.018	0.021	0.018	0.007	0.002	0.002
RMSE	0.135	0.105	0.120	0.104	0.041	0.010	0.010
SMAPE	2.927	2.056	2.384	2.007	1.067	0.305	0.294

at different time scales are calculated, and draw the q-DSCID results of different models at different time scales τ with error bar in Fig. 7. It is shown that when the initial scale $\tau = 1$, the q-DSCID value of Type I is the smallest in all the models, which is obviously superior to other models. With the increase of time scale τ, the q-DSCID value gap between the models gradually decrease and tend to stabilize. The CID value of Type I for WTI, BRE, HO, and NG is kept to a minimum, and the value of NG is very close to that of Type II. It can be considered that the prediction result of Type I is optimal for WTI, BRE, HO, and NG. The Type II prediction of NG is also very good. Combining the basic metrics and complex metrics of each model, we can conclude that the RIF-DBGRUNN model is the most accurate and stable among comparison models.

The proposed RIF-DBGRUNN is optimized based on GRU. To better reflect the property improvement after model optimization, combined with Tables 4–7, the model accuracy metrics increase percentage graph based on GRU accuracy is drawn in Fig. 8. The way increase percentage (IP) is constructed can be denoted as

$$\mathrm{IP} = \begin{cases} \dfrac{Metrics_{Others} - Metrics_{GRU}}{Metrics_{GRU}} & Others = \mathrm{R}^2 \\ -\dfrac{Metrics_{Others} - Metrics_{GRU}}{Metrics_{GRU}} & Otherwise \end{cases} \tag{26}$$

where $Metrics_{GRU}$ represents the accuracy metrics of GRU, and $Metrics_{Others}$represents the accuracy metrics of other models.

Fig. 8 shows a histogram of the metrics increase ratio for WTI, BRE, HO and NG. It can be found that in the prediction results of the four data sets, the prediction accuracy of the GRU model is better than that of the ERNN model and the SVM model, because their IP values are negative, and the GRU and LSTM models are not much

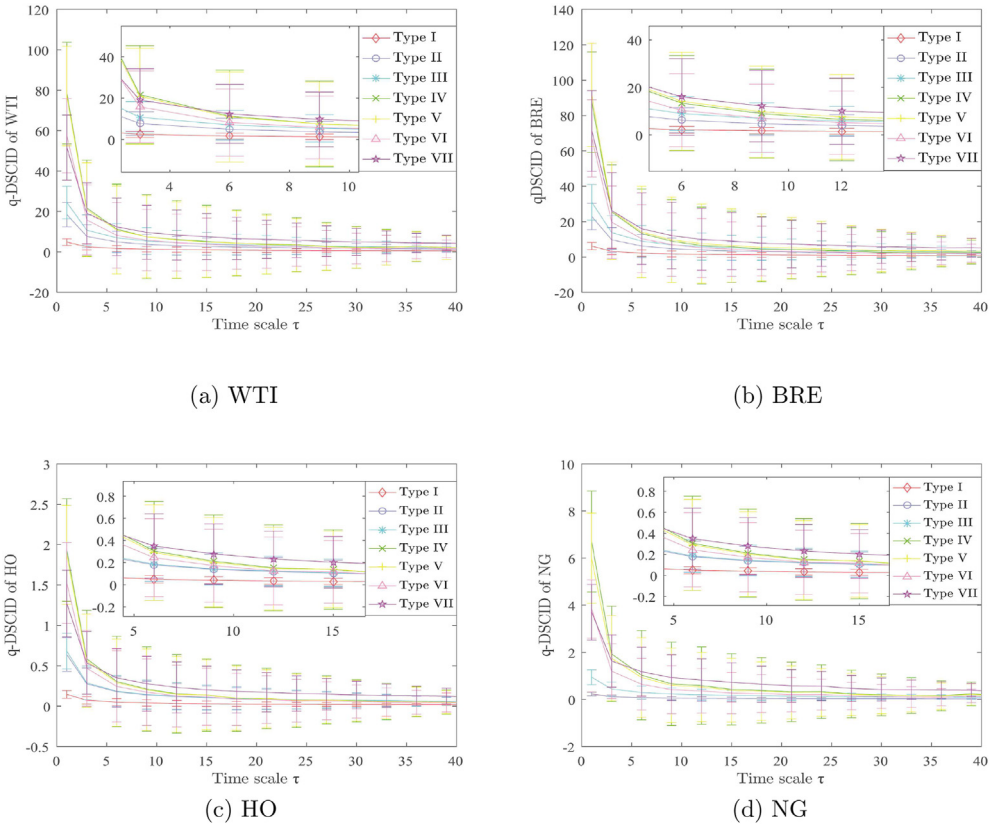

Fig. 7. Variation of q-DSCID values of each model on scale τ.

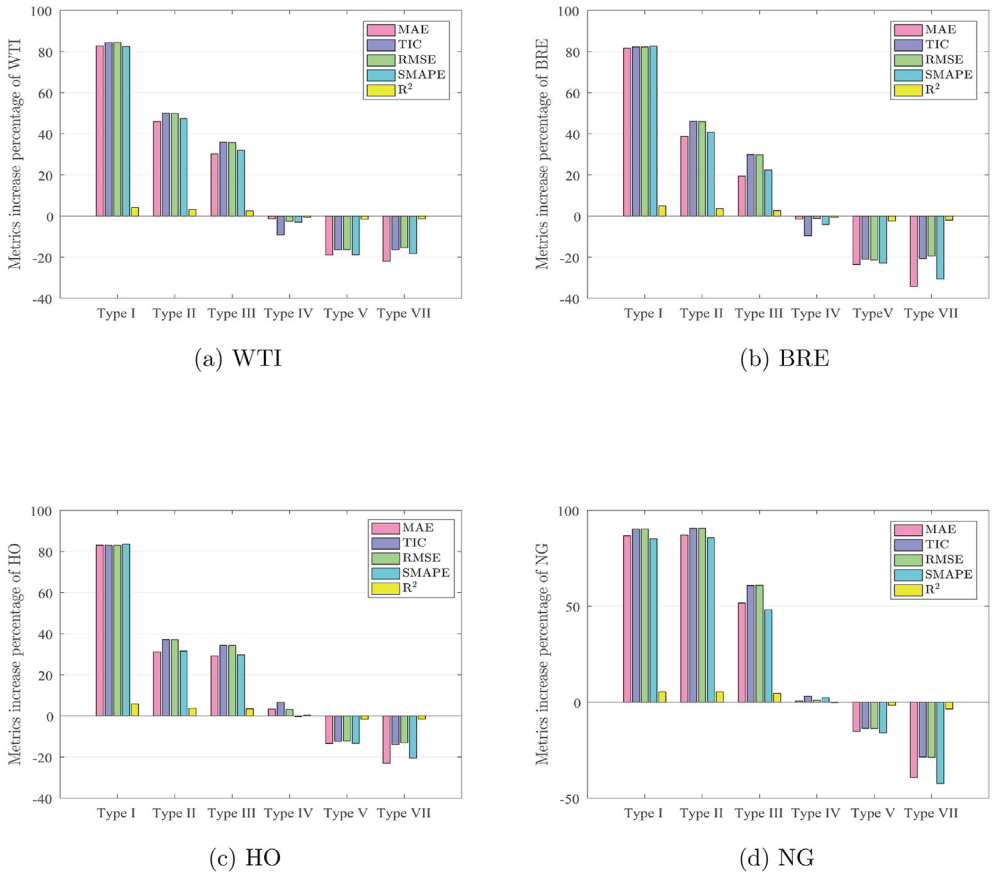

Fig. 8. Percent histogram of metrics improvement based on GRU accuracy.

different. The LSTM model's prediction results for WTI and BRE are not as good as GRU model, but it is better in the prediction of HO and NG. Moreover, The IP values of the improved DBGRUNN, RIF-GRUNN and RIF-DBGRUNN models are all positive, which means that the results are better than the GRU model, especially the RIF-DBGRUNN model. Table 8 shows the specific values of the increase percentage of each model. DBGRUNN is built on the basis of GRU by integrating the deep bidirectional learning structure, and Type III is obtained. Taking the WTI data set as an example, it can be seen from Table 8 that the prediction accuracy of DBGRUNN has been improved, and the indexes of MAE, TIC, RMSE and SMAPE have increased by more than 30% to 30.167%, 35.992%, 35.804%, and 32.024%, respectively. Among them, because the base of R^2 is close to the upper limit of 1, the increase is relatively small at 2.491%. On the basis of GRU, combined with the RIF-based training method, Type II is obtained. The increase rates of MAE, TIC, RMSE and SMAPE indexes exceeds 40%, reaching 45.996%, 50.133%, 50.070%, and 47.459%, and R^2 increased by 3.172%. Type I is obtained based on Type III combined with RIF-based training method, and its accuracy indexes have been improved significantly. Among them, the indicators of MAE, TIC, RMSE and SMAPE are increased by more than 80%compared with GRU, which are 82.811%, 84.337%, 84.362%, and 82.474%with R^2 improved 4.12%. The prediction results of WTI demonstrate that constructing deep bidirectional network and applying RIF-based training algorithm can effectively improve prediction accuracy. Analyzing the IP valus of WTI, BRE, HO, and NG prediction results of each model, it can be conclude that the performance of the novel RIF-DBGRUNN model outperforms the comparison models.

Crude oil futures have a pivotal significance in real life. It can increase the country's pricing influence in the international oil market. Especially for oil-consuming countries, the price fluctuations of crude oil futures can affect the country's energy security.

Table 8
Increase percentage of each model relative to the GRU model for metrics.

	Model					
	Type I	Type II	Type III	Type IV	Type V	Type VII
WTI						
MAE	82.811	45.996	30.167	−1.273	−18.915	−21.954
TIC	84.337	50.133	35.992	−9.151	−16.332	−16.296
RMSE	84.362	50.070	35.804	−2.529	−16.283	−15.331
SMAPE	82.474	47.459	32.049	−3.006	−18.789	−18.204
R^2	4.120	3.172	2.491	−0.540	−1.489	−1.315
BRE						
MAE	81.618	38.808	19.386	−1.460	−23.582	−34.252
TIC	82.306	46.042	29.988	−9.648	−21.035	−20.620
RMSE	82.317	45.962	29.746	−1.202	−21.389	−19.458
SMAPE	82.609	40.773	22.414	−4.065	−22.829	−30.658
R^2	4.927	3.605	2.596	−0.662	−2.358	−2.017
HO						
MAE	83.068	31.109	29.155	3.274	−13.371	−22.998
TIC	82.926	37.163	34.304	6.581	−12.282	−13.973
RMSE	82.939	37.077	34.240	3.155	−12.158	−13.066
SMAPE	83.659	31.557	29.697	−0.283	−13.315	−20.543
R^2	5.886	3.672	3.448	0.346	−1.575	−1.501
NG						
MAE	87.146	86.727	51.687	0.665	−15.221	−39.241
TIC	90.613	90.225	60.861	3.198	−13.660	−28.579
RMSE	90.600	90.211	60.938	1.042	−13.714	−28.699
SMAPE	85.698	85.143	48.100	2.348	−15.962	−42.361
R^2	5.386	5.382	4.606	−0.260	−1.593	−3.562

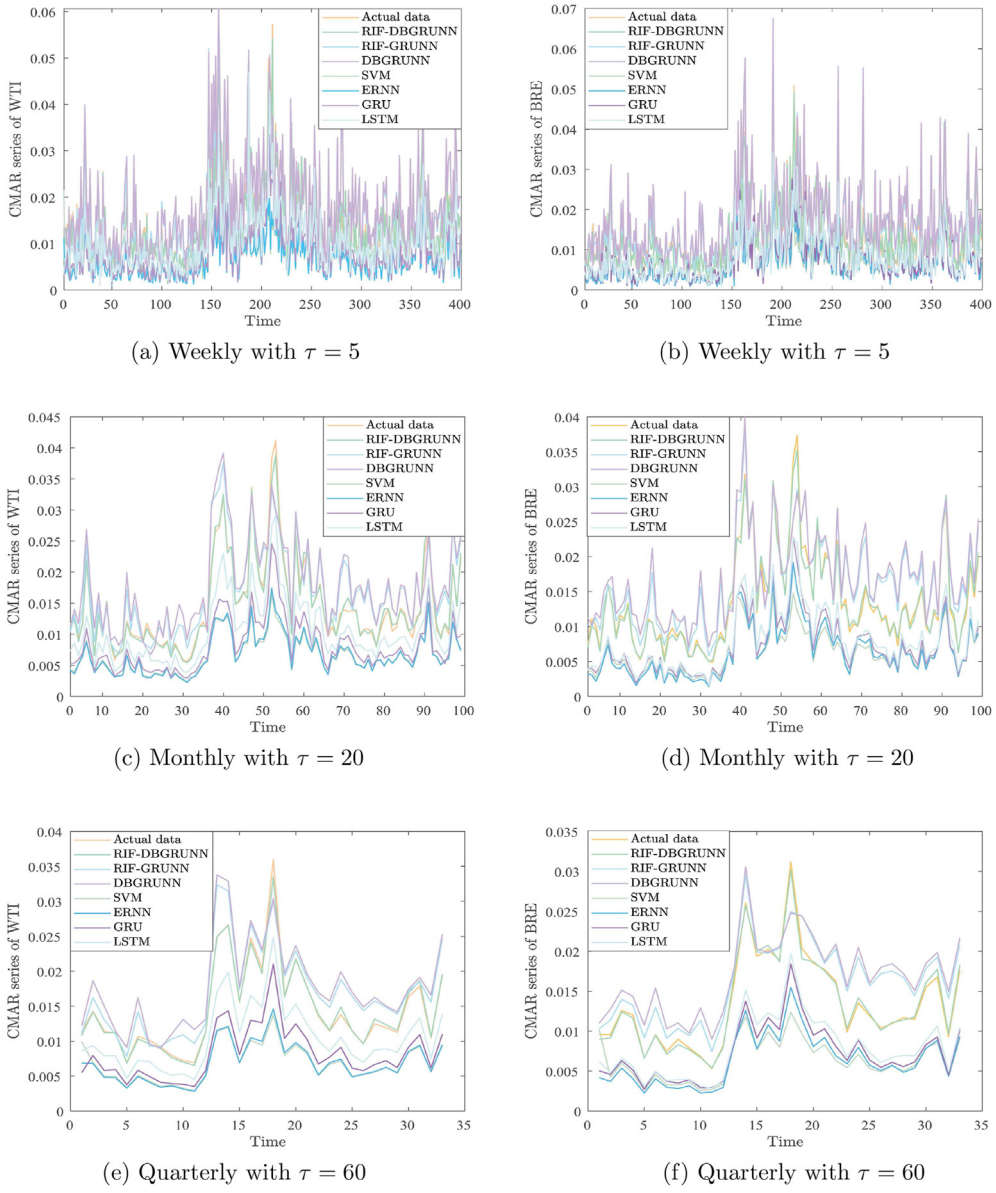

Fig. 9. Prediction results of weekly, monthly, and quarterly coarse-grained moving absolute return series for WTI and BRE by different models. (a)(c)(e) for WTI. (b)(d)(f) for BRE.

The futures market has the function of discovering prices, and the crude oil futures trading market can reflect the expectations, demands and production costs of the market's micro entities, so it has a price-oriented function. Moreover, the crude oil futures market can provide crude oil consuming enterprises with hedging and risk aversion channels. If a more reliable prediction of crude oil futures prices can be achieved, it will play a positive role in the stable production and operation of oil-related companies. Crude oil futures are a tool for relevant companies to hedge their risks, and can escort the stable operation of companies and even the entire energy market. The price forecast of crude oil futures can provide a theoretical basis for the government departments to supervise the crude oil futures market and macro-control. It can also provide a reference for crude oil companies and individual participants to better use crude oil futures to avoid operating risks.

4.4. Model validation of coarse-grained moving absolute return series

The return series is an important goal of financial research. It fully reflects the investment opportunities of the asset, is easier to handle than the price series, and has better statistical properties. An excellent logarithmic return is given below, expressed as

$$r(t) = \ln\left(1 + \frac{\mathscr{P}(t) - \mathscr{P}(t-1)}{\mathscr{P}(t-1)}\right) = \ln\left(\frac{\mathscr{P}(t)}{\mathscr{P}(t-1)}\right) = \ln(\mathscr{P}(t)) - \ln(\mathscr{P}(t-1)) \tag{27}$$

where $\mathscr{P}(t)$ represents the closing price at time t.

Given a random set of price return series $r = (r_1, r_1, \cdots, r_n)$, define a novel coarse-grained moving absolute return series $R^{(\tau)} = (R_1^{(\tau)},$

$R_2^{(\tau)}, \cdots, R_{\frac{n}{\tau}}^{(\tau)}$), where $\left\lfloor \frac{n}{\tau} \right\rfloor$ means rounding down to $\frac{n}{\tau}$. The point of each coarse-grained moving absolute return (CMAR) series is expressed as

$$R_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} |r(i)|, \quad 1 \le j \le \frac{n}{\tau}. \tag{28}$$

When $\tau = 1$, it is the initial price return rate series, we consider the case of $\tau = 5, 20, 60$ respectively corresponding to weekly, monthly, and quarterly coarse-grained moving absolute return series. Simultaneously, it corresponds to the needs of the majority of short-term, medium-term and long-term investors.

The proposed model is applied to predict the series of CMAR and compared with other models. The WTI and BRE data sets in Table 2 are selected for the experiment, and the prediction results are shown in Fig. 9. Fig. 9(a)(c)(e) and (b)(d)(f) respectively display the prediction results and actual data of the CMAR series of each model when $\tau = 5, 20, 60$. It can be seen that no matter WTI or BRE, the prediction results of CMAR series by RIF-DBGRUNN under any τ scales are the closest to actual data, indicating that RIF-DBGRUNN model also has a good prediction performance for the CMAR series.

Table 9 lists the accuracy metrics of the CMAR series prediction for each model when τ is 5, 20, and 60 respectively. At $\tau = 5$, the R^2, MAE, TIC, RMSE, and SMAPE index values of RIF-DBGRUNN model for WTI are 0.979, 0.001, 0.037, 0.001, and 6.692, which are significantly better than other comparison models. Similarly, it can be found from the table that the results produced by the RIF-DBGRUNN model have higher accuracy. Simultaneously, as the value of τ increases, the number of samples decreases but the accuracy of the model's prediction results generally improves. The increase of τ makes the series smoother and the prediction result also improves. Experiments show that the model's prediction result for the series with large fluctuations is not as good as the prediction result with small fluctuations. It also means that the proposed model will be more credible in the medium and long term prediction results.

Table 9
Metrics of coarse-grained moving absolute return series forecast results with different τ scales.

Model	Metrics				
	R^2	MAE	TIC	RMSE	SMAPE
	WTI				
$\tau = 5$					
RIF-DBGRUNN	0.979	0.001	0.037	0.001	6.692
RIF-GRUNN	0.647	0.004	0.149	0.005	27.743
DBGRUNN	0.541	0.005	0.170	0.006	32.557
SVM	0.205	0.008	0.422	0.011	75.135
ERNN	0.231	0.008	0.413	0.010	75.075
GRU	0.392	0.007	0.320	0.009	60.027
LSTM	0.612	0.005	0.209	0.006	36.680
$\tau = 20$					
RIF-DBGRUNN	0.988	0.001	0.025	0.001	4.025
RIF-GRUNN	0.730	0.003	0.112	0.004	20.424
DBGRUNN	0.639	0.004	0.132	0.005	25.878
SVM	0.245	0.008	0.389	0.009	73.347
ERNN	0.270	0.008	0.380	0.009	73.199
GRU	0.425	0.007	0.288	0.007	57.724
LSTM	0.649	0.004	0.177	0.005	34.305
$\tau = 60$					
RIF-DBGRUNN	0.992	0.000	0.019	0.001	2.453
RIF-GRUNN	0.768	0.003	0.098	0.003	18.737
DBGRUNN	0.679	0.004	0.118	0.004	23.533
SVM	0.255	0.008	0.381	0.009	73.544
ERNN	0.278	0.008	0.372	0.009	73.155
GRU	0.435	0.006	0.280	0.007	57.524
LSTM	0.652	0.004	0.171	0.005	34.169
	BRE				
$\tau = 5$					
RIF-DBGRUNN	0.977	0.001	0.040	0.001	7.897
RIF-GRUNN	0.561	0.004	0.169	0.006	30.030
DBGRUNN	0.456	0.005	0.189	0.008	34.200
SVM	0.221	0.008	0.424	0.010	75.224
ERNN	0.254	0.008	0.401	0.010	76.930
GRU	0.362	0.007	0.339	0.008	65.382
LSTM	0.481	0.006	0.287	0.007	57.822
$\tau = 20$					
RIF-DBGRUNN	0.988	0.001	0.024	0.001	3.994
RIF-GRUNN	0.675	0.003	0.125	0.004	23.405
DBGRUNN	0.578	0.004	0.146	0.005	28.685
SVM	0.276	0.008	0.386	0.009	75.257
ERNN	0.323	0.007	0.360	0.008	75.956
GRU	0.402	0.006	0.305	0.007	63.786
LSTM	0.485	0.006	0.263	0.006	56.483
$\tau = 60$					
RIF-DBGRUNN	0.992	0.000	0.017	0.001	2.911
RIF-GRUNN	0.680	0.003	0.112	0.004	22.901
DBGRUNN	0.582	0.004	0.132	0.004	28.171
SVM	0.283	0.008	0.378	0.008	74.898
ERNN	0.332	0.007	0.351	0.008	74.946
GRU	0.411	0.006	0.296	0.007	62.998
LSTM	0.490	0.006	0.257	0.006	55.717

5. Conclusion

In order to predict the energy market futures price more accurately, a novel deep bidirectional GRU neural network model RIF-DBGRUNN based on the random inheritance formula is proposed. It integrates the method of deep learning and further ameliorates the prediction accuracy of the model through bidirectional learning structure and RIF-based training algorithm. The verification results signify that the proposed model stands out among comparison models with high prediction accuracy.

Applying the RIF-DBGRUNN to predict WTI, BRE, HO, and NG data sets respectively, the prediction results are compared with real data. The *RE*results illustrate that the prediction results are excellent. Besides, the predicted results are compared with six different models SVM, GRU, ERNN, LSTM, DBGRUNN and RIF-GRUNN, and the values of MAE, TIC, RMSE, SMAPE and R^2 indicate that the prediction results of the RIF-DBGRUNN model are superior to other comparison models, which are 0.175, 0.002, 0.214, 0.298, and 0.999 for WTI. The consequence of linear regression analysis denotes that the prediction of RIF-DBGRUNN has trivial error with the true value, and most of the continuous minimum values of the q-DSCID synchronization evaluation further verify that the prediction accuracy of the model is higher. Moreover, the analysis of IP indicator shows the validity of the random inheritance formula and the deep bidirectional learning structure. In addition, the experimental results reveal that the proposed model also has a good performance in predicting the coarse-grained moving absolute return series, and it is concluded that the model has a better prediction ability for smoother series, which can provide a more reliable reference for medium and long term investors to make decisions.

Generally speaking, the novel RIF-DBGRUNN model performs well in energy market futures price forecasting, and further research can consider the following aspects. (i) Explore the selection of relevant parameters in the random inheritance formula, such as drift function and volatility function. (ii) Consider combining feature extraction methods to build new hybrid network models, such as empirical mode decomposition, variational mode decomposition. (iii) Consider extending the model to other application areas.

Credit author statement

Bin Wang: Conceptualization; Methodology; Investigation;

Software; Visualization; Writing – original draft and review & editing. Jun Wang: Conceptualization; Methodology; Investigation; Supervision; Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors were supported by National Natural Science Foundation of China Grant No. 71271026.

References

[1] Hart R. Non-renewable resources in the long run. J Econ Dynam Contr 2016;71:1–20.
[2] Szoplik J. Forecasting of natural gas consumption with artificial neural networks. Energy 2015;85. 208-20.
[3] Gong X, Lin B. Time-varying effects of oil supply and demand shocks on China's macro-economy. Energy 2018;149. 424-37.
[4] Niu H, Wang J. Return volatility duration analysis of NYMEX energy futures and spot. Energy 2017;140. 837-49.
[5] Kristjanpoller W, Minutolo MC. Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Syst Appl 2016;65. 233-41.
[6] Cheng FZ, Li T, Wei YM, Wei FTJ. The vec-nar model for short-term forecasting of oil prices. Energy Econ 2019;78. 656-67. itemfactor1 Yurtsever C, Zahor T. Oil Price Shocks and Stock Market in the Netherlands. Working Paper. University of Groningen; 2007.
[7] Ederington LH, Fernando CS, Hoelscher SA, Lee TK, Linn SC. A review of the evidence on the relation between crude oil prices and petroleum product prices. Journal of Commodity Markets 2019;13:1–15.
[8] Yu Y, Wang J. Lattice-oriented percolation system applied to volatility behavior of stock market. J Appl Stat 2012;39. 785-97.
[9] Jw E, Bao YL, Ye JM. Crude oil price analysis and forecasting based on variational mode decomposition and independent component analysis. Phys Stat Mech Appl 2017;485. 412-27.
[10] Jammazi R, Aloui C. Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Econ 2012;34. 828-41.
[11] Safari A, Davallou M. Oil price forecasting using a hybrid model. Energy 2018;148:49–58.
[12] Abiyev RH. Fuzzy wavelet neural network based on fuzzy clustering and gradient techniques for time series prediction. Neural Comput Appl 2011;20. 249-59.
[13] Solgi Y, Ganjefar S. Variable structure fuzzy wavelet neural network controller for complex nonlinear systems. Appl Soft Comput 2018;64. 674-85.
[14] Fu S, Li Y, Sun S, Li H. Evolutionary support vector machine for RMB exchange rate forecasting. Phys Stat Mech Appl 2019;521. 692-704.
[15] Li Y, Che J, Yang Y. Subsampled support vector regression ensemble for short term electric load forecasting. Energy 2018;164. 160-70.
[16] Guo XP, Li DC, Zhang AH. Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters. AASRI Procedia 2012;1. 525-30.
[17] Kao LJ, Chiu CC, Lu CJ, Yang JL. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. Neurocomputing 2013;99. 534-42.
[18] Hong YY, Liu CY, Chen SJ, Huang WC, Yu TH. Short-term LMP forecasting using an artificial neural network incorporating empirical mode decomposition. International Transactions on Electrical Energy Systems 2015;25. 1952-64.
[19] Kordanuli B, Barjaktarovic L, Jeremic L, Alizamir M. Appraisal of artificial neural network for forecasting of economic parameters. Phys Stat Mech Appl 2017;465. 515-9.
[20] Kusumo F, Silitonga AS, Masjuki HH, Ong HC, Siswantoro J, Mahlia TMI. Optimization of transesterification process for Ceiba pentandra oil: a comparative study between kernel-based extreme learning machine and artificial neural networks. Energy 2017;134:24–34.
[21] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy 2010;35. 512-7.
[22] Keles D, Scelle J, Paraschiv F, Fichtner W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl Energy 2016;162. 218-30.
[23] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Energy 2018;212. 372-85.
[24] Pan Y, Zheng RC, Zhang JX, Yao X. Predicting bike sharing demand using recurrent neural networks. Procedia Computer Science 2019;147. 562-6.
[25] Hajiabotorabi Z, Kazemi A, Samavati FF, Ghaini FMM. Improving DWT-RNN model via B-spline wavelet multiresolution to forecast a high-frequency time series. Expert Syst Appl 2019;138:112842.
[26] Berradi Z, Lazaar M. Integration of principal component analysis and recurrent neural network to forecast the stock price of Casablanca stock exchange. Procedia Computer Science 2019;148. 55-61.
[27] Wang J, Wang J. Forecasting energy market indices with recurrent neural networks: case study of crude oil price fluctuations. Energy 2016;102. 365-74.
[28] Graves A. Supervised sequence labelling with recurrent neural networks. Springer Berlin Heidelberg; 2012.
[29] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9. 1735-80.
[30] Wu W, Liao W, Miao J, Du G. Using gated recurrent unit network to forecast short-term load considering impact of electricity price. Energy Procedia 2019;158. 3369-74.
[31] Chen J, Jing H, Chang Y, Liu Q. Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process. Reliab Eng Syst Saf 2019;185. 372-82.
[32] Hooman A, Seyed BE. A new hybrid model for forecasting Brent crude oil price. Energy 2020;200:117520.
[33] Qiao WB, Yang Z. Forecast the electricity price of U.S. using a wavelet transform-based hybrid model. Energy 2020;193:116704.
[34] Cen ZP, Wang J. Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. Energy 2019;169. 160-71.
[35] Jw E, Ye JM, He LL, Jin HH. Energy price prediction based on independent component analysis and gated recurrent unit neural network. Energy 2019;189:116278.
[36] Donate JP, Cortez P, Sanchez GG, Miguel AS. Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble. Neurocomputing 2013;109:27-32.
[37] Jammazi R, Aloui C. Crude oil forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Econ 2012;34. 828-41.
[38] Niu HL, Wang J. Volatility clustering and long memory of financial time series and financial price model. Digit Signal Process 2013;23. 489-98.
[39] Gatheral J, Schied A. Optimal trade execution under geometric Brownian motion in the almgren and chriss framework. Int J Theor Appl Finance 2011;14. 353-68.
[40] Dufresne D. The integral of geometric brownian motion. Adv Appl Probab 2001;33. 223-41.
[41] Wu SD, Wu CW, Lin SG, Wang CC, Lee KY. Time series analysis using composite multiscale entropy. Entropy 2013;15. 1069-84.
[42] Niu HL, Wang J. Quantifying complexity of financial short-term time series by composite multiscale entropy measure. Commun Nonlinear Sci Numer Simulat 2015;22. 375-82.
[43] Greff K, Srivastava RK, Koutnik J, Steunebrink BR. LSTM: a search space odyssey. IEEE Transactions on Neural Networks & Learning Systems 2015;28. 2222-32.
[44] J E, Ye J, He L, Jin H. Energy price prediction based on independent component analysis and gated recurrent unit neural network. Energy 2019;189:116278.
[45] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436.
[46] Tumer AE, Kocer S, Koca A. Estimation of the electricity consumption of Turkey trough artificial neural networks. IEEE International Symposium on Computational Intelligence & Informatics 2017.
[47] Movagharnejad K, Mehdizadeh B, Banihashemi M, Kordkheili MS. Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. Energy 2011;36. 3979-84.
[48] Milacic L, Jovic S, Vujovic T, Miljkovic J. Application of artificial neural network with extreme learning machine for economic growth estimation. Phys Stat Mech Appl 2017;465. 285-8.
[49] Liu H, Tian HQ, Liang XF, Li YF. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157. 183-94.
[50] Ruiz LGB, Rueda R, Cuellar MP, Pegalajar MC. Energy consumption forecasting based on elman neural networks with evolutive optimization. Expert Syst Appl 2018;92. 380-9.