Engineering Applications of Artificial Intelligence 47 (2016) 110–121

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting

Lean Yu, Wei Dai, Ling Tang*

School of Economics and Management, Beijing University of Chemical Technology, 15 Beisanhuan East Road, Beijing 100029, China

ARTICLE INFO

Available online 28 May 2015

Keywords:
Crude oil price forecasting
New production development
Artificial intelligence
Decomposition-and-ensemble learning paradigm
Extended extreme learning machine

ABSTRACT

As one of the most important energy resources, an accurate prediction for crude oil price can effectively guarantee a rapid new production development with higher production quality and less production cost. Accordingly, a novel decomposition-and-ensemble learning paradigm integrating ensemble empirical mode decomposition (EEMD) and extended extreme learning machine (EELM) is proposed for crude oil price forecasting, based on the principle of "decomposition and ensemble". This novel learning model makes contribution to literature by introducing the current powerful artificial intelligent (AI) technique of EELM in the ensemble model formulation. In the proposed method, EEMD, a competitive decomposition method, is first applied to divide the original data of crude oil price time series into a number of relatively regular components, for simplicity. Second, EELM, a currently proposed, powerful, effective and stable forecasting tool, is implemented to predict all components independently. Finally, these predicted results are aggregated into an ensemble result as final prediction, using simple addition ensemble method. For illustration and verification purposes, the proposed learning paradigm is used to predict the crude oil spot price of WTI. Empirical results demonstrate that the proposed novel ensemble learning paradigm statistically outperforms all considered benchmark models (including popular single models and similar ensemble models) in both prediction accuracy (in terms of level and directional measurement) and effectiveness (in terms of time saving and robustness), indicating that it is a promising tool to predict complicated time series with high volatility and irregularity.

1. Introduction

International crude oil prediction has become an increasingly hot issue within the research fields of energy analysis and economic management, which can effectively guarantee a rapid new production development with higher production quality and less production cost. First, due to its significant role in the global economy and society (Oman, 2003), an accurate prediction for crude oil market is extremely indispensable for stable and rapid economic development and thence new production development. In particular, a leap in crude oil price would result in an inflation and economy recession in oil-consuming nations, and further negatively impact global economy. In contrast, a fast falling of crude oil price would otherwise prohibit the economic development of oil-producing countries, and further generate political instability and social unrest (Gholamian et al., 2005; Chen and Hsu, 2012). Therefore, an accurate prediction for crude oil price can effectively help capture the market dynamics and make the corresponding policies for avoiding high volatility of crude oil price and thus reducing the market risk, which can further enable a stable macroeconomic environment for a rapid new production development. Second, as one of the most important energy inputs, an accurate prediction for crude oil price can effectively help make appropriate production plans for new products in terms of higher quality and less cost. In particular, a higher crude oil price may enhance the production cost with the same use of crude oil, and vice versa. Therefore, an accurate prediction for crude oil price can effectively help make and revise production plans of new products and techniques for determining various inputs, which can significantly enhance the quality and reduce the cost of new production development. However, it has been proved to be an extremely tough task of forecasting crude oil price, due to the interactive inner factors, such as supply and demand, competition across providers, substitution with other energy forms, technique development, domestic economy, deregulation activities, globalization and even uncertainties caused by political instabilities, wars and conflicts (Chen, 2009; He et al., 2012; Zhang et al., 2009). To address such tough task, this paper concentrates on crude oil price forecasting, in order to improve the prediction performance from both prediction accuracy and time saving perspectives.

According to existing literature, a variety of forecasting models have been formulated for international crude oil price prediction.

* Corresponding author. Tel./fax: +86 10 6441 2210.
E-mail address: tangling@mail.buct.edu.cn (L. Tang).

http://dx.doi.org/10.1016/j.engappai.2015.04.016
0952-1976/

Generally, there are two main categories for the crude oil price forecasting. The first category can be referred to traditional statistical and econometric techniques, such as linear regression (LinR), generalized auto regressive conditional heteroskedasticity (GARCH) family models, random walk (RW), grey model (GM) and error correction models (ECM). For example, a sophisticated econometric model was applied to predict crude oil price (Huntington, 1994). Lin (2009) predicted the international crude oil futures price via GM(1,1). Hou and Suardi (2012) implemented a nonparametric GARCH model to predict the return volatility in oil price. Mohammadi and Su (2010) proposed a novel hybrid model, coupling ARIMA and GARCH models, to estimate the conditional mean and volatility of weekly crude oil spot prices in eleven international markets. Similarly, Murat and Tokat (2009) employed RW to forecast oil price movements. Lanza et al. (2005) investigated the prices of crude oil and oil products by using ECM model. Besides, autoregressive integrated moving average (ARIMA) model, the most typical traditional time series model, has also frequently been applied as benchmark in crude oil price forecasting (e.g., He et al., 2012; Yu et al., 2008; Li et al., 2013).

However, these above traditional econometric techniques may be insufficient to capture the hidden nonlinear features in crude oil price (Bao et al., 2007; Yu et al., 2007), and there is a need to find a new approach to remedy the shortcomings of the traditional methods. In the previous studies, artificial intelligence (AI) models with powerful self-learning capacities, such as artificial neural networks (ANNs), support vector machine (SVM) and other intelligent optimization algorithms, have become increasingly popular for crude oil price forecasting recently, and the empirical results demonstrated their superiority to traditional methods. For ANN, Abdullah and Zeng (2010) introduced ANN to analyze the quantitative data of crude oil price. Kulkarni and Haidar (2009) presented a multilayer feed-forward neural network (FNN) to predict crude oil spot price. Kaboudan (2001) employed genetic programming (GP) and ANN to forecast crude oil price. As far as SVM, Xie et al. (2006) implemented SVM model for crude oil price forecasting and compared its prediction performance with ARIMA and back-propagation neural network (BPNN). Khashman and Nwulu (2011) employed SVM to predict crude oil price. Li and Ge (2013) improved ε-support vector regression (ε-SVR) machine with dynamic errors correction for crude oil price forecasting. All these studies demonstrated that the AI models are quite superior to the statistical-based models in modeling the nonlinear and complicated data of crude oil price.

Though the AI models (e.g., ANN and SVM) are very effective relative to traditional models, AI models also have their own shortcomings. For example, the time wasting, slow convergence and local minima may be the most important disadvantages, especially in ANN. In order to overcome these drawbacks, a novel learning algorithm called extreme learning machine (ELM), a special case of single hidden layer feedforward networks (SLFNs) proposed by Huang et al. (2004), tends to provide a better generalization performance and much faster learning speed than the above gradient learning algorithms, without setting stopping criteria, learning rate and learning epochs.

According to existing literature, there was few research about crude oil price forecasting by using ELM, although ELM has wildly been implemented in other forecasting cases, such as medium-term sales in fashion retail supply chains (Wong and Guo, 2010; Sun et al., 2008), electricity prices (Shrivastava and Panigrahi, 2014; Tian and Meng, 2010) and other applications (Wang and Han, 2015), and the empirical results all witnessed that ELM significantly outperformed its counterparts (e.g., ARIMA, SVR and ANN models) in both level and directional forecasting (Liu et al., 2012; Pati et al., 2013). Since ELM might be somewhat unstable with randomicity (Sun et al., 2008; Rong et al., 2008; Singh and Balasundaram, 2007; Miche et al., 2010), an extended ELM (EELM) method was accordingly proposed (Sun et al., 2007), where a given number of ELM models are run and the average value of the prediction results is calculated as the final result. Since EELM might be more stable and accurate than its original form, EELM model is especially introduced here as a very promising approach for forecasting international crude oil price.

Besides, a "decomposition and ensemble" principle can be also considered as a helpful tool for analyzing the data with high complexity and irregularity (Yu et al., 2008; Wang et al., 2005). Actually, the effectiveness of "decomposition and ensemble" has been already confirmed, and a series of decomposition-and-ensemble learning paradigms have been accordingly proposed. For instance, Yu et al. (2008) proposed a novel empirical mode decomposition (EMD) based neural network ensemble learning paradigm to predict the crude oil price. Tang et al. (2011) selected ensemble EMD (EEMD) and least squares support vector regression (LSSVR) respectively as decomposition and forecasting tools, to formulate a EEMD-based LSSVR learning paradigm for forecasting nuclear power consumption. Wang et al. (2014) integrated the EMD and Elman neural network to predict the wind speed. Lu and Shao (2012) put forward an ensemble approach integrating EEMD and ELM for forecasting computer products sales. Wang et al. (2011) proposed a seasonal decomposition (SD) based LSSVR learning approach for hydropower consumption forecasting. Tang et al. (2015) built a novel decomposition ensemble model by coupling the complementary EEMD and EELM, for crude oil price forecasting. Yu et al. (2014) constructed a similar methodology based on compressed sensing (CS) as data decomposition technique and some powerful AI forecasting tools, for crude oil price forecasting. All empirical results statistically verified that the methodology framework of "decomposition and ensemble" can significantly improve prediction performance. Therefore, this study tends to conduct the prediction research for international crude oil price under such effective "decomposition-and-ensemble" model framework.

Generally speaking, based on the "decomposition and ensemble" principle, this study tries to propose a novel "decomposition-and-ensemble" learning paradigm integrating EEMD and EELM, i.e., EEMD-based EELM ensemble learning paradigm, to forecast the international crude oil price. In this proposed methodology, the original data of crude oil price time series are first divided into several relatively independent intrinsic mode functions (IMFs) and one residue by EEMD, an efficient decomposition method relative to other decomposition methods (e.g., EMD and wavelet decomposition). Second, EELM, a fast and powerful forecasting tool relative to traditional statistical techniques and other AI models (e.g., ANN and SVM techniques), is applied to predict the different IMFs and residue independently. Finally, these predicted values are fused into an ensemble result as the final prediction by simple addition (ADD) ensemble method, since the sum of real values of the decomposed components is actually equal to the original data. The main contribution of the paper is to introduce the current powerful AI technique of EELM in the decomposition-and-ensemble method formulation. Different from other existing decomposition-and-ensemble models, this novel method especially utilizes the currently proposed EELM technique as the individual forecasting tool, with its unique merits of powerful prediction capability, time-saving training process and model robustness.

The main motivation of this study is to formulate a novel EEMD-based EELM ensemble learning paradigm to improve the performance of international crude oil price prediction, in terms of prediction accuracy, time saving, and robustness, and to compare its prediction performance with other popularly used forecasting techniques (including typical single models and similar ensemble models). The rest of this study is organized as follows. Section 2

describes the formulation process of the proposed EEMD-based EELM ensemble learning paradigm in detail. For illustration and verification purposes, crude oil spot price of West Texas Intermediate (WTI) is used to test the effectiveness of the proposed methodology in multi-step-ahead predictions, as the experiment study designed in Section 3. The corresponding results and effectiveness of the proposed method are discussed in Section 4. Finally, some concluding remarks and future researches are drawn in Section 5.

2. Methodology formulation

Generally speaking, there are three main steps involved in the proposed decomposition-and-ensemble methodology, i.e., decomposition, individual forecast and ensemble forecast. In this section, the overall formulation process of the EEMD-based EELM ensemble learning paradigm is presented. First, EEMD and EELM algorithms are briefly introduced in Sections 2.1 and 2.2, respectively. Then, the EEMD-based EELM ensemble learning algorithm is formulated and the corresponding steps are described in Section 2.3.

2.1. Ensemble empirical mode decomposition (EEMD)

Ensemble empirical mode decomposition (EEMD) is extended from empirical model decomposition (EMD) to overcome the intrinsic drawback of mode mixing (Wu and Huang, 2009). Unlike other traditional decomposition methodologies, such as Fourier decomposition and wavelet decomposition, EEMD and EMD are empirical, intuitive, direct and self-adaptive data processing methods, designed especially for nonlinear and nonstationary data (Zhang et al., 2010). Since the seminal works of Wu and Huang, (2009) and Huang et al. (1998) were published, EMD and EEMD have widely been applied to various complex system analyses, and the results indicated the effectiveness of EMD and EEMD in decomposing complex time series data (Zhang et al., 2008; Xie et al., 2008).

The main purpose of EEMD is to extract the coexisting oscillatory functions, namely intrinsic mode functions (IMFs), from the original time series. Mathematically, IMFs must satisfy the following two conditions: (1) in each whole function, the number of extrema (i.e., both maxima and minima) and number of zero crossings must be equal, or different at the most by one, and (2) the functions must be symmetric with respect to local zero mean. Hence, the original data series x_t,(t=1,2,…,T) can be expressed as a sum of IMFs and one residue:

$$x_t = \sum_{j=1}^{N} c_{j,t} + r_{N,t} \tag{1}$$

where N is the number of IMFs, $r_{N,t}$ is the final residue, and $c_{j,t}$, $(j = 1, 2, ..., N)$ is the jth IMF, at time t. The IMF components containing in different frequency bands are different and change with variation of time series x_t, while $r_{N,t}$ represents the central tendency of data series x_t. In practice, the total number of IMFs, N, can be determined based on the data size T, according to $N = \log_2 T$ (Wu and Huang, 2009).

Different from EMD, EEMD involves an additional step of adding white noise, under the assumption that the observed data are amalgamations of true time series and noise, and the ensemble means of data with different noise levels are closer to true time series. Under such assumption, the white noise series can help extract the true IMFs, and offset themselves via ensemble averaging. The effect of the added white noise can be controlled according to the well-established statistical rule (Wu and Huang, 2009):

$$\varepsilon_{ne} = \frac{\varepsilon}{\sqrt{NE}} \tag{2}$$

where NE is the number of ensemble members, ε is the amplitude of the added noise, and ε_{ne} is the final standard deviation of error, defined as the difference between original signal and the corresponding IMFs.

2.2. Extended extreme learning machine (EELM)

Since the extended extreme learning machine (EELM) is an improved version of extreme learning machine (ELM), this subsection first introduces the ELM algorithm and then EELM.

The capability of modified SLFNs (i.e., ELM) has been already verified by Huang et al. (2006a). For standard SLFNs with $\tilde{N}$ hidden neuron nodes and activation function $g(x)$, given a dataset with N arbitrary distinct samples $(\mathrm{x}_i,\mathrm{t}_i)$, i=1,2,…,N (where $\mathrm{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,n}]^T \in R^n$ is the input data and $\mathrm{t}_i = [t_{i,1}, t_{i,2}, ..., t_{i,m}]^T \in R^m$ is the target output), the mathematical model of ELM can be formulated as (Huang et al., 2006b)

$$\sum_{i=1\tilde{N}} \beta_i g_i(\mathrm{x}_j) = \sum_{i=1\tilde{N}} \beta_i g_i(\mathrm{w}_i \cdot \mathrm{x}_j + b_i) = \mathrm{t}_j, \quad j = 1, 2, ..., N \tag{3}$$

where $\mathbf{w}_i = [w_{i,1}, w_{i,2}, ..., w_{i,n}]^T$ denotes the weight vector which connects the input nodes with the ith hidden neuron node, $\beta_i = [\beta_{i,1}, \beta_{i,2}, ..., \beta_{i,m}]^T$ is the weight vector which connects the output nodes with the ith hidden neuron node, and b_i is the threshold of the ith hidden neuron node. The inner product of $\mathbf{w}_i$ and $\mathbf{x}_j$ is denoted by the operation $\mathbf{w}_i \cdot \mathbf{x}_j$.

In order to make the standard SLFNs with $\tilde{N}$ hidden neuron nodes and activation function $g(x)$ approximate to these samples with zero error, the SLFNs' input weights $\mathbf{w}_i$ and the hidden layer biases b_i are specified by training an SLFN (the input weights w_i and hidden biases b_i are generated randomly in ELM-based NNs), and this train process can be seen as finding a minimum norm least squares (LS) solution $\hat{\beta}$ for the linear system $\mathbf{H}\beta = \mathbf{T}$, where $\mathbf{H}$ denotes the hidden layer output matrix, and T indicates the matrix of targets.

$$\mathbf{H} = \begin{bmatrix} h(\mathbf{x}_1) \\ \cdots \\ h(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_{\tilde{N}}(\mathbf{x}_1) \\ \cdots & \cdots & \cdots \\ h_1(\mathbf{x}_N) & \cdots & h_{\tilde{N}}(\mathbf{x}_N) \end{bmatrix} \tag{4}$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T & \cdots & \cdots & \mathbf{t}_N^T \end{bmatrix}^T \tag{5}$$

Accordingly, the evaluation of the output weights linking the hidden layer to the output layer is equivalent to exploring the LS solution to the given linear system:

$$\|\mathbf{H}(\mathbf{w}_1, ..., \mathbf{w}_{\tilde{N}}, \boldsymbol{b}_1, ..., \boldsymbol{b}_{\tilde{N}})\hat{\beta} - \mathbf{T}\| = \min_{\beta} \|\mathbf{H}(\mathbf{w}_1, ..., \mathbf{w}_{\tilde{N}}, \boldsymbol{b}_1, ..., \boldsymbol{b}_{\tilde{N}})\beta - \mathbf{T}\| \tag{6}$$

where $\hat{\beta} = \begin{bmatrix} \beta_1^T & \cdots & \cdots & \beta_{\tilde{N}}^T \end{bmatrix}^T$ is the matrix of output weights. Therefore, the minimum norm LS solution to the linear system can be described in Eq. (7):

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \tag{7}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose (MP) generalized inverse of matrix $\mathbf{H}$ (Rao and Mitra, 1971). The minimum norm LS solution is unique and has the smallest norm among all LS solutions.

Since ELM might be somewhat unstable with random values of input weights $\mathbf{w}$ and hidden layer biases b (Sun et al., 2008; Rong et al., 2008; Singh and Balasundaram, 2007), an extended ELM (EELM) was accordingly proposed where we run the original ELM for P times and calculate the average values of prediction as final prediction results, as illustrated in Fig. 1.

In Fig. 1, $\mathbf{t}_j^k, (k=1,2,\ldots,P)$ is the prediction value via the kth ELM and $\bar{\mathbf{t}}_j = 1/P \sum_{k=1}^{P} \mathbf{t}_j^k$ is the final forecasting result. Obviously, the larger the parameter P is, the more stable the final results $\bar{\mathbf{t}}_j$ might be. In particular, EELM is generally composed of the following four main steps:

(1) Randomly initialize the input weights and hidden layer biases for the kth ELM, i.e., $\mathbf{w}_i^k$ and b_i^k, respectively.
(2) Single ELM model is performed to obtain a prediction result $\mathbf{t}_j^k$ for $\mathbf{x}_j$, with the parameters $\mathbf{w}_i^k$ and b_i^k.
(3) Repeat steps (1) and (2) with different random $\mathbf{w}_i^k$ and b_i^k for P times (i.e., $k=1,2,\ldots,P$), and thus generate P different prediction results $\mathbf{t}_j^k$.
(4) Compute final prediction result by averaging all single prediction results, i.e., $\bar{\mathbf{t}}_j = 1/P \sum_{k=1}^{P} \mathbf{t}_j^k$.

2.3. EEMD-based EELM ensemble learning paradigm

Based on the general framework of "decomposition-and-ensemble" methodology (Yu et al., 2008; Tang et al., 2012) and the previous techniques, a three-step ensemble learning paradigm integrating ensemble empirical mode decomposition (EEMD) and extended extreme learning machine (EELM) can be formulated for time series data forecasting.

As shown in Fig. 2, the proposed EEMD-based EELM ensemble learning paradigm is generally composed of the following three main steps:

(1) The original time series $x_t, (t=1,2,\ldots,T)$ is decomposed into N intrinsic mode function (IMF) components $c_{j,t}, (j=1,2,\ldots,N)$ and one residue component $r_{N,t}$ by using EEMD algorithm.
(2) EELM is used as individual forecasting tool to model the extracted IMF components and the residue component independently. Accordingly, the prediction results $\hat{c}_{j,t}$ and $\hat{r}_{N,t}$ for the corresponding components can be obtained.
(3) Prediction results of all IMF components and the residue component are summed up into the final results $\hat{x}_t$ for original data x_t, by using ADD approach as the ensemble tool.

To summarize, the proposed EEMD-based EELM ensemble learning paradigm can be abbreviated as an "EEMD (decomposition) – EELM (individual prediction) – ADD (ensemble prediction)" ensemble learning approach, based on the principle of "decomposition and ensemble".

3. Experimental design

In order to verify the effectiveness of the proposed decomposition-and-ensemble learning paradigm, the crude oil spot price of West Texas Intermediate (WTI) is selected as experimental sample, as mentioned in Section 3.1. Section 3.2 gives the main evaluation criteria for prediction capability.

3.1. Data description

The crude oil spot price data of West Texas Intermediate (WTI) are obtained from the US energy information administration (EIA) (http://www.eia.doe.gov/). The sample data are daily data covering the period from January 2, 1986 to October 21, 2013 with a total of 7015 observations. In this study, the sample data are divided into two subsets: training subset and testing subset. We treat the data from January 2, 1986 to March 31, 2008 as training set, including 5612 observations used for model training, while the data from April 1, 2008 to October 21, 2013 (1403 observations) as the testing set used to evaluate prediction performance. It is worth noting that multi-step-ahead predictions at horizons one, three and six

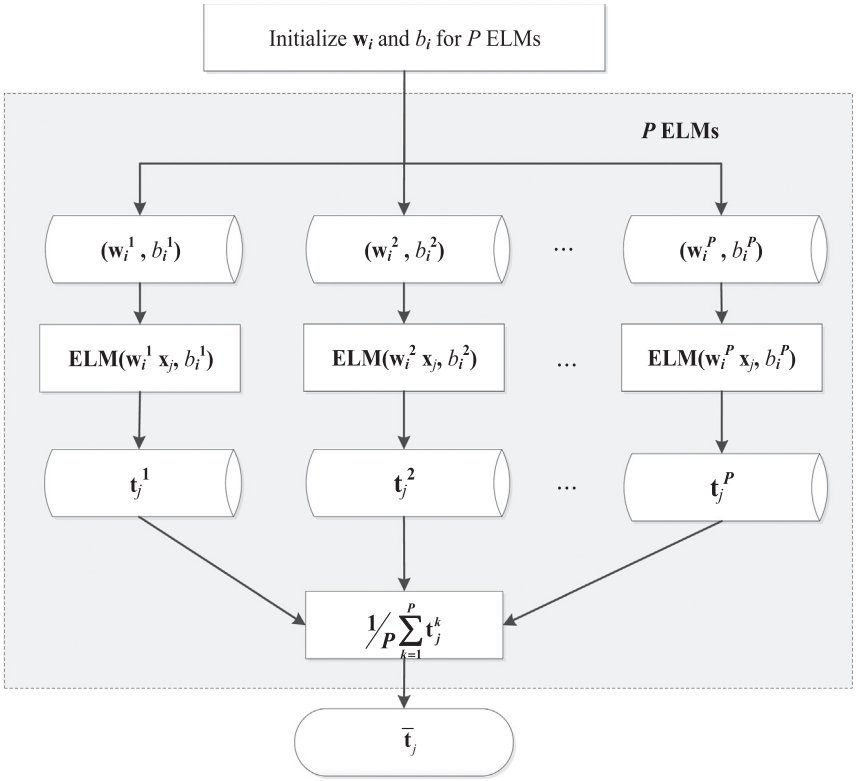

Fig. 1. The framework of the EELM technique.

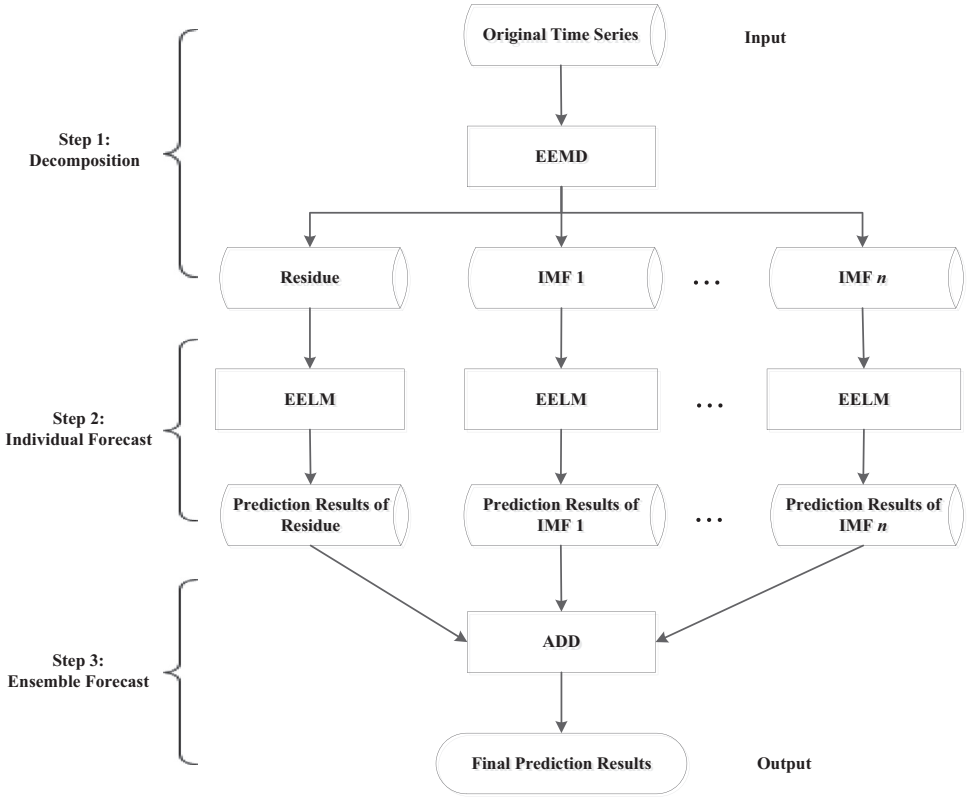

Fig. 2. Overall process of the EEMD-based EELM ensemble learning paradigm.

are performed in this study (Liu et al., 2013). Given a time series $x_t, (t = 1, 2, ..., T)$, we make m-step-ahead prediction for x_{t+m}:

$$\hat{x}_{t+m} = f(x_t, x_{t-1}, ..., x_{t-(l-1)}) \tag{8}$$

where $\hat{x}_{t+m}$ is the m-step-ahead predicted value at time t, x_t is the actual value at time t, and l presents the lag orders. The lag orders l is set through autocorrelation and partial correlation analysis in this study (Yu et al., 2014).

3.2. Evaluation criteria

First, the root mean squared error (*RMSE*) and the mean absolute percent error (*MAPE*) are selected to evaluate level prediction accuracy (Wang et al., 2011).

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right|, \quad RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} \tag{9}$$

where N is the number of observations in the testing dataset.

Furthermore, the capability of predicting direction of movement can be measured by a directional statistic (D_{stat}) (Yu et al., 2014):

$$D_{stat} = \frac{1}{N}\sum_{t=1}^{N} a_t \times 100\% \tag{10}$$

where $a_t = 1$ if $(\hat{x}_{t+1} - x_t)(x_{t+1} - x_t) \geq 0$, or otherwise $a_t = 0$.

In order to further compare different forecasting models from statistical perspective, Diebold–Mariano (DM) statistic is used to test the statistical significance of the prediction difference (Diebold and Mariano, 1995). In this study, the loss function is set to mean square prediction error (*MSPE*) and the null hypothesis is that the *MSPE* of the tested model *A* is not less than that of the benchmark model *B*. In particular, the DM statistic can be defined as

$$S = \frac{\bar{g}}{\left(\hat{V}_{\bar{g}}/N\right)^{1/2}} \tag{11}$$

where $\bar{g} = 1/N\sum_{t=1}^{N} g_t$, $g_t = (x_t - \hat{x}_{A,t})^2 - (x_t - \hat{x}_{B,t})^2$ and $\hat{V}_{\bar{g}} = \gamma_0 + 2\sum_{l=1}^{\infty}\gamma_l, (\gamma_l = \mathrm{cov}(g_t, g_{t-1}))$. $\hat{x}_{A,t}$ and $\hat{x}_{B,t}$ represent the predicted values for x_t respectively calculated by the target method A and benchmark method B, at time t. Here, unilateral test is used to test the S statistic.

4. Experimental results

The detailed steps of experimental study, together with the parameter specification, are first given in Section 4.1, and the corresponding results are further discussed in Section 4.2.

4.1. Parameter specification

For comparison purpose, some other popular forecasting models are employed as benchmarks to compare with the proposed ensemble forecasting approach in prediction performance.

Based on previous literature investigation, the most popular univariate forecasting models, especially for crude oil price series, are introduced, i.e., the most typical traditional statistical model of ARIMA (Box and Jenkins, 1970; Deng, 1982), and the most popular AI techniques of ANN (Wu and Chau, 2010; Yu et al., 2009) and LSSVR (Xie et al., 2013; Quan et al., 2010). In particular, these three tools of ARIMA, ANN and LSSVR might be the most popular time series forecasting models, which have widely been used as single prediction models (e.g., Tang et al., 2014) and individual prediction model of decomposition-and-ensemble methodologies (e.g., Yu et al., 2014). Therefore, these three typical univariate forecasting tools are especially introduced as single benchmark models in this

study. Furthermore, since EELM is selected as the individual forecasting tool for extracted components in formulating the proposed novel learning paradigm, we also introduce EELM and its original form ELM as single benchmark models.

In the case of decomposition, EMD, the original form of EEMD, is also considered as benchmark decomposition model. As for individual prediction tools for the extracted components, the four powerful AI techniques, i.e., EELM, ELM, ANN and LSSVR, can also be implemented. Similarly, they can be also utilized as ensemble prediction approaches, to compare with the ADD ensemble approach in the proposed ensemble learning model.

Generally speaking, to verify the prediction performance of the proposed learning paradigm, i.e., EEMD-EELM-ADD, five single models, (i.e., single ARIMA, EELM, ELM, ANN and LSSVR models) and eleven hybrid learning paradigms (i.e., EEMD-EELM-EELM, EEMD-LSSVR-LSSVR, EEMD-LSSVR-ADD, EEMD-ANN-ANN, EEMD-ANN-ADD, EMD-EELM-EELM, EMD-EELM-ADD, EMD-LSSVR-LSSVR, EMD-LSSVR-ADD, EMD-ANN-ANN and EMD-ANN-ADD) are formulated as benchmark methods in this study.

First, five single models, i.e., EELM, ELM, LSSVR, ANN and ARIMA, are performed to forecast crude oil price of WTI as the benchmark models. In single EELM and ELM methods, the number of hidden neurons is set to 30, by trial-and-error method (Sun et al., 2007), and the sigmoidal function, i.e., $g(x)=1/1+e^{-x}$, is chosen as the activation function. For EELM, the number of ensemble members P is set to 100 (Wong and Guo, 2010). For single LSSVR model, Gaussian RBF kernel function is selected, and values of parameters sigma squared and gamma are set by trial-and-error method (Tang et al., 2012). In single ANN (I–H–O) model, hidden nodes H is set to 7, output neuron O is one and input neurons I is determined by autocorrelation and partial correlation analyses. All ANN models are iteratively run 10,000 times to train the model, using the training subset (Yu et al., 2014).The parameters of ARIMA (p–d–q) is estimated by Schwarz criterion (SC) minimization (Yu et al., 2008).

Second, twelve ensemble learning paradigms are performed. In EEMD, the added white noise has a standard deviation of 0.2 and the ensemble member is set to 100. Fig. 3 shows the decomposition results of original crude oil price data via EEMD, where all IMF components are listed in the order from the highest frequency to the lowest frequency, and the last one is the residue. The second and third steps of the ensemble learning paradigms are individual and ensemble forecasting, via different individual and ensemble tools. For consistency purpose, all AI models (i.e., EELM, ELM, ANN and LSSVR) as both individual and ensemble prediction tools, follow a similar way to those of respective single benchmarks, as mentioned above.

4.2. Results analysis

For a clear discussion, four steps are involved in results analysis. First, performances of the five single methods are compared with each other to select the best single model. Subsequently, the results of the proposed EEMD-EELM-ADD model, other ensemble benchmarks and the best single model are discussed, in order to make sure whether the proposed ensemble learning paradigm can be statistically proved to be superior to all considered benchmarks in terms of prediction accuracy and time saving. Furthermore, the robustness of the novel learning paradigm is analyzed. Finally, some interesting conclusions can be obtained from the experimental study.

4.2.1. Performance comparison of single methods

The prediction performances of five single models (i.e., EELM, ELM, LSSVR, ANN and ARIMA) are discussed in this section. The prediction accuracy is evaluated via *MAPE*, *RMSE* and D_{stat}, as respectively shown in Figs. 4–6, while Fig. 7 reports the comparison results for model effectiveness in terms of consuming time. From these results, one conclusion can be obviously drawn that the EELM model might be the most powerful single model in terms of prediction accuracy in most cases, with much less computing time.

In terms of level accuracy (measured by *MAPE* and *RMSE*), it is easy to find that prediction accuracy of EELM is the highest amongst the five single benchmark models, followed by other AI models, and the ARIMA model ranks the last. First, when comparing AI models with

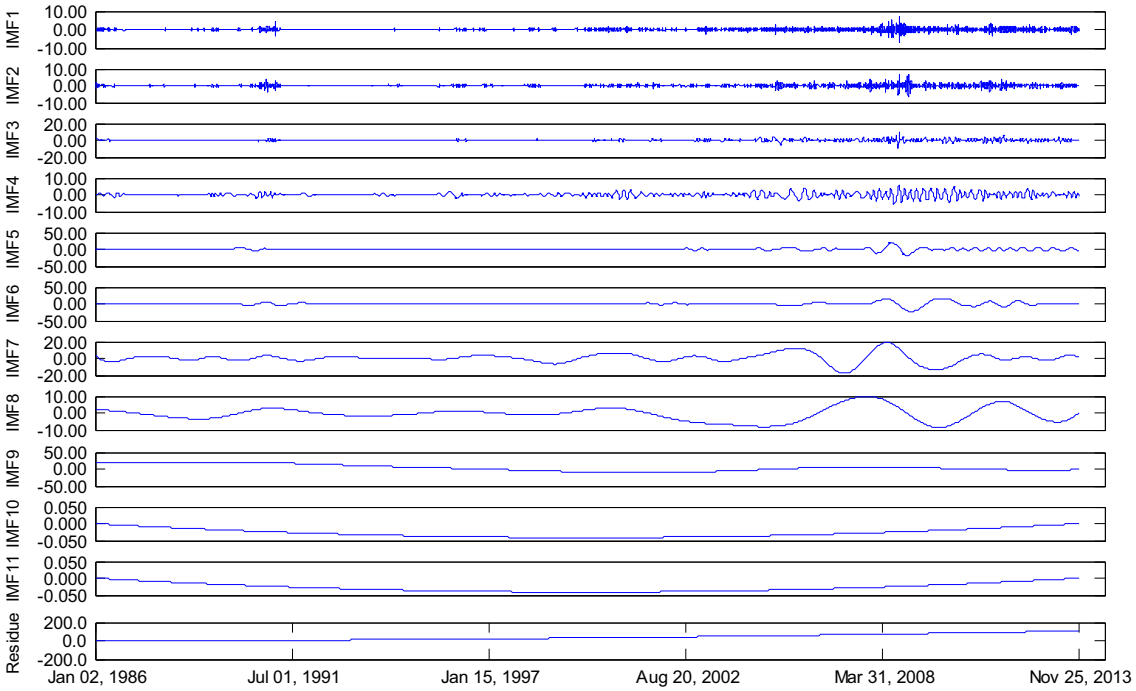

Fig. 3. The IMFs and one residue for international crude oil price via EEMD.

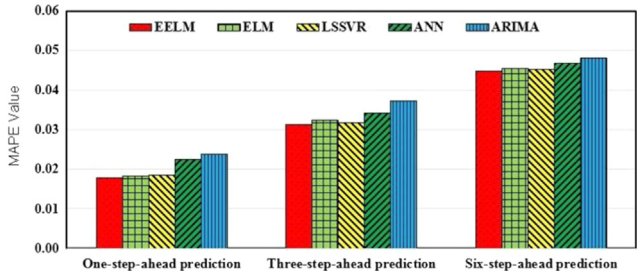

Fig. 4. Performance comparison of different single methods in terms of *MAPE* criterion.

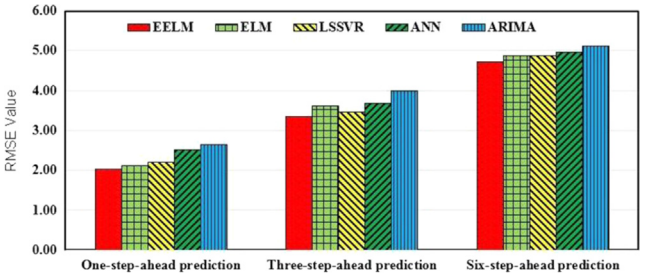

Fig. 5. Performance comparison of different single methods in terms of *RMSE* criterion.

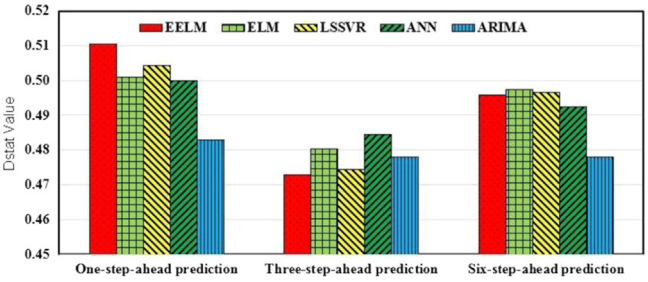

Fig. 6. Performance comparison of different single methods in terms of D_{stat} criterion.

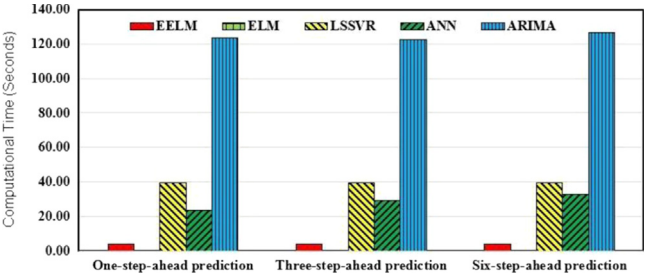

Fig. 7. Performance comparison of different single methods in terms of computational time (seconds).

traditional model, it can be obviously found that all AI models, i.e., EELM, ELM, LSSVR and ANN, produce much better prediction results than ARIMA in all cases, in terms of *MAPE* and *RMSE*. The possible reason is that due to the nonlinearity and complexity of crude oil price data, ARIMA as a typical linear model might find difficulty, while AI tools are more suitable for capturing the nonlinear patterns. Second, as for AI models, EELM outperforms all other AI models measured by *MAPE* and *RMSE* in all cases, while the prediction performances of other three models are quite alike. The results further imply the superiority of EELM to other popular AI tools with their intrinsic weakness, such as local minima and parameter sensitiveness. Third, while the prediction results by ELM are quite unstable, ranking the second at horizon one and the third at horizons three and six in terms of *MAPE* and *RMSE*, EELM can be shown to be much more stable and robust, ranking the first in all cases. The main reason may be referred to the model improvement by EELM through averaging prediction results obtained by a set of ELMs, which effectively addresses the problem of randomicity in ELM.

Focusing on directional accuracy, two important conclusions can be found. First, it can be obviously seen that all AI techniques can produce much higher directional prediction accuracy in terms of D_{stat}, while the traditional ARIMA model ranks the last in most cases. The hidden reason may be referred to the effectiveness of AIs in modeling nonlinear, irregular and complex data, especially the crude oil price data. Second, as for the AI models, their directional prediction performances are quite alike, and none of them can be fully proved to be better than the others, even though EELM has been shown to be the best in level prediction accuracy. The main reason might be the limitation of single models, while the concept of "decomposition and ensemble" can be introduced to helpfully improve the prediction performance of models.

From the perspective of time saving, Fig. 7 obviously shows that EELM and ELM are the most effective models, with much less computing time than other single benchmarks (i.e., LSSVR, ANN and ARIMA). This demonstrates the effectiveness of EELM and ELM, especially in addressing the problems of time wasting and slow convergence in typical AI techniques. When comparing EELM with ELM, ELM is undoubtedly faster than EELM, since a total of 100 ELMs are run in each EELM method.

In order to statistically test the differences amongst the five single benchmark models in prediction capability, the DM test is used, as the results reported in Tables 1–3. First, it can be obviously seen that all *p*-values of EELM over other benchmarks are far below 0.1, except the comparison results between EELM and ELM, indicating the superiority of EELM. Second, when EELM and ELM are tested against other single models, all *p*-values are below 0.1 (see the first two lines in Tables 1–3), indicating the superiority of EELM and ELM over other AI techniques and traditional ARIMA model at the confidence level of 90%. Third, when the traditional ARIMA model is tested as benchmark model, all *p*-values are far below 0.1 (see the last columns in Tables 1–3), confirming the effectiveness of AI models in modeling nonlinear patterns in crude oil price data, relative to traditional models. Fourth, when comparing EELM with ELM, the difference in prediction performance seems not significant, since the *p*-values are all above 0.1. The main reason may be that even though EELM is extended from ELM to address the instability weakness, the superiority of EELM to ELM might not be significant in prediction accuracy. Finally, when comparing LSSVR with ANN, none of them can be statistically proved to be better than the other. This implies that, there is no obvious difference between LSSVR and ANN when they are used to predict the crude oil price.

4.2.2. *Performance comparison of decomposition-and-ensemble learning paradigms*

Focusing on ensemble learning paradigms (i.e., EEMD-EELM-EELM, EEMD-EELM-ADD, EEMD-LSSVR-LSSVR, EEMD-LSSVR-ADD,

Table 1
DM test results for single models in one-step-ahead prediction.

Tested model	Benchmark model			
	ELM	LSSVR	ANN	ARIMA
EELM	−0.15 (0.4403)	−2.87 (0.0021)	−4.68 (0.0000)	−12.15 (0.0000)
ELM		−2.34 (0.0095)	−4.94 (0.0000)	−12.32 (0.0000)
LSSVR			−0.17 (0.4337)	−7.43 (0.0000)
ANN				−9.58 (0.0000)

EEMD-ANN-ANN, EEMD-ANN-ADD, EMD-EELM-EELM, EMD-EELM-ADD, EMD-LSSVR-LSSVR, EMD-LSSVR-ADD, EMD-ANN-ANN and EMD-ANN-ADD), Figs. 8–11 show the comparison results in terms of *MAPE*, *RMSE*, D_{stat} and computing time, respectively, and Tables 4–9 report the results of DM test with different horizons. Furthermore, the best single model, EELM model, is also mentioned for comparison.

From these results, one important conclusion can be obviously seen that the proposed model of EEMD-EELM-ADD is measured to be the best model in both prediction accuracy and time saving for crude oil price, and the DM test statistically confirms its superiority at the confidence level of 90%.

In terms of level accuracy measured by *MAPE* and *RMSE* (see Figs. 8 and 9), some interesting findings can be summarized. First, the proposed EEMD-EELM-ADD outperforms all benchmarks in all cases, implying that EEMD-EELM-ADD is the best learning paradigm for crude oil price forecasting in terms of level prediction accuracy. In contrast, the poorest models might be EMD-LSSVR-LSSVR and EEMD-LSSVR-LSSVR, even much worse than single EELM model. The main reason might be referred to the intrinsic weakness of LSSVR, i.e., parameter sensitiveness. Second, focusing on decomposition algorithms in these ensemble methods, it can be found that the ensemble learning algorithms with EEMD as decomposition method are much better than their counterpart methods with EMD in terms of level prediction accuracy. This indicates that EEMD is a more effective decomposition method than EMD in data analysis. Third, as for individual forecasting tools for extracted components, EEMD-EELM-ADD significantly outperforms the EEMD-ANN-ADD and EEMD-LSSVR-ADD learning paradigms without exception, even though the latter two models consume much more computing time (see Fig. 11). This repeatedly confirms the superiority of EELM to other AI tools in prediction. Fourth, in terms of ensemble tools, it can be found that all models with ADD ensemble method (i.e., EEMD-EELM-ADD, EEMD-ANN-ADD, EEMD-LSSVR-ADD, EMD-EELM-ADD, EMD-ANN-ADD and EMD-LSSVR-ADD) outperform their respective counterparts with the related AIs as ensemble tools. Specially, EMD-EELM-EELM and EEMD-LSSVR-LSSVR obtain quite poor performances. The possible reason may be that under the effective strategy of "decomposition and ensemble", the ADD ensemble method may be sufficient to generate a satisfactory final result, while AI ensemble models might confront over-fitting problem. Therefore, it can be concluded from the results that ADD is a simple but effective ensemble technique for formulating ensemble learning paradigms. Finally, the EEMD-EELM-ADD learning paradigm is significantly better than its single EELM model. This implies that the decomposition strategy does effectively improve prediction performance of model.

Focusing on D_{stat} criterion (shown in Fig. 10), the proposed EEMD-EELM-ADD learning paradigm performs much better than other models with the highest directional hits in all cases, indicating its powerful prediction capability. First, all ensemble learning models are shown to be much better in directional prediction, while the single EELM model achieves the lowest value of D_{stat}, which implies that the ensemble learning paradigms with decomposition strategy are much effective in directional prediction relative to the single models. Second, amongst the twelve ensemble learning methods, the EEMD-LSSVR-LSSVR is the poorest model in terms of directional accuracy. This again implies that LSSVR, without a careful parameters optimization process which consumes lots of time in large dataset case, might lose power in both individual and ensemble prediction. Finally, it is not hard to find that the hit-rates of all methods tend to decrease as the prediction horizon increases, and the main hidden reason may be referred to the information loss in a multi-step-ahead prediction with a large horizon.

When comparing the consuming time of these different models (shown in Fig. 11), three conclusions can be drawn. First, amongst ensemble learning paradigms, the models with EELM or ELM as individual (or both individual and ensemble) forecasting tools are much effective than other ensemble models, in terms of time saving, indicating the improvement of EELM and ELM over other AI tools in fast convergence. This further implies the effectiveness of the proposed EEMD-EELM-ADD, in terms of both prediction accuracy and time saving. Second, when comparing ensemble models with the single EELM model, the single EELM model might be much faster, and the hidden reason is easy to understand that all steps of decomposition, individual forecast for all extracted components and ensemble forecast in these ensemble methods need to expend extra computing time. Third, focusing on decomposition tools, EMD-based AI ensemble models consume less computing time than their counterparts with EEMD as decomposition algorithm, and the reason may be referred to

Table 2
DM test results for single models in three-step-ahead prediction.

Tested model	Benchmark model			
	ELM	LSSVR	ANN	ARIMA
EELM	−0.60 (0.2730)	−2.13 (0.0167)	−2.82 (0.0024)	−8.90 (0.0000)
ELM		−2.22 (0.0131)	−2.14 (0.0161)	−8.68 (0.0000)
LSSVR			−0.32 (0.3758)	−6.31 (0.0000)
ANN				−7.09 (0.0000)

Table 3
DM test results for single models in six-step-ahead prediction.

Tested model	Benchmark model			
	ELM	LSSVR	ANN	ARIMA
EELM	−0.44 (0.3300)	−1.36 (0.0863)	−1.75 (0.0397)	−6.69 (0.0000)
ELM		−1.26 (0.1043)	−1.82 (0.0340)	−6.99 (0.0000)
LSSVR			−0.04 (0.4844)	−2.45 (0.0072)
ANN				−4.55 (0.0000)

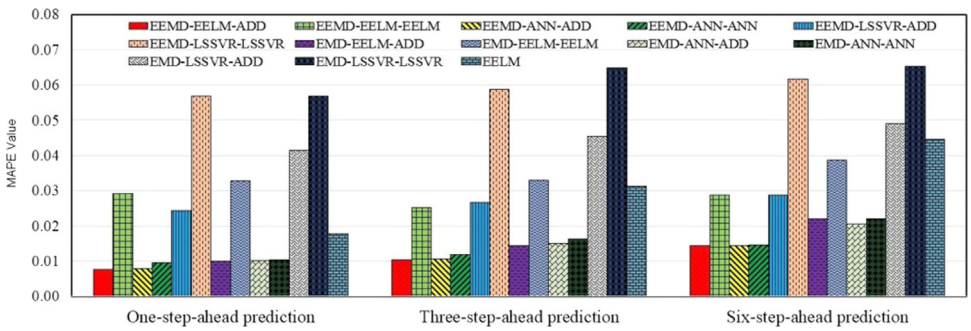

Fig. 8. Performance comparison of different ensemble learning methods and EELM in terms of *MAPE* criterion.

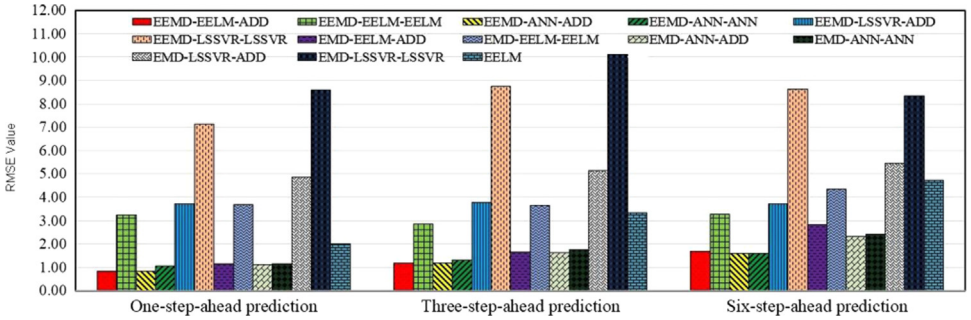

Fig. 9. Performance comparison of different ensemble learning methods and EELM in terms of *RMSE* criterion.

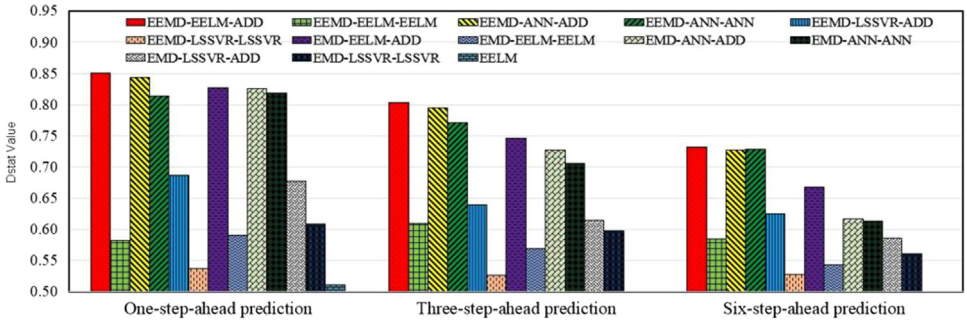

Fig. 10. Performance comparison of different ensemble learning methods and EELM in terms of D_{stat} criterion.

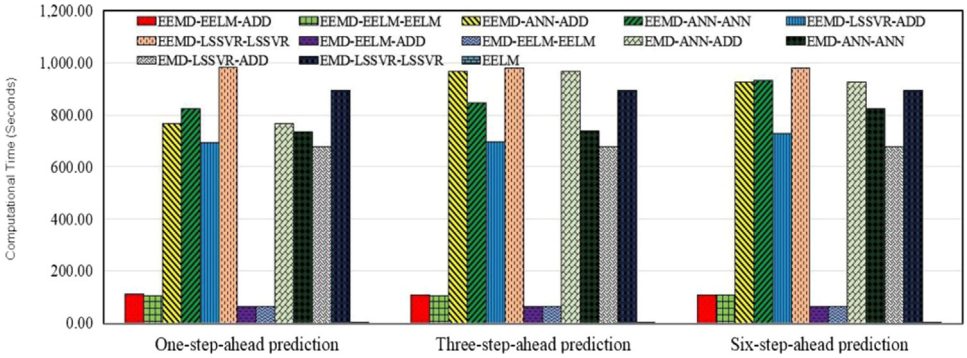

Fig. 11. Performance comparison of different ensemble learning methods and EELM in terms of computational time (seconds).

Table 4
DM test results for EEMD/EMD-EELM-ADD/EELM and EELM models in one-step-ahead prediction.

Tested model	Benchmark model			
	EEMD-EELM-EELM	**EMD-EELM-ADD**	**EMD-EELM-EELM**	**EELM**
EEMD-EELM-ADD	−19.43 (0.0000)	−14.17 (0.0000)	−17.73 (0.0000)	−24.59 (0.0000)
EEMD-EELM-EELM		16.69 (1.0000)	−2.99 (0.0014)	10.35 (1.0000)
EMD-EELM-ADD			−15.31 (0.0000)	−20.80 (0.0000)
EMD-EELM-EELM				10.19 (1.0000)

Table 5
DM test results for EEMD/EMD-EELM-ADD/EELM and EELM models in three-step-ahead prediction.

Tested model	Benchmark model			
	EEMD-EELM-EELM	**EMD-EELM-ADD**	**EMD-EELM-EELM**	**EELM**
EEMD-EELM-ADD	−13.87 (0.0000)	−10.86 (0.0000)	−16.58 (0.0000)	−21.37 (0.0000)
EEMD-EELM-EELM		9.73 (1.0000)	−6.88 (0.0000)	−3.94 (0.0000)
EMD-EELM-ADD			−13.54 (0.0000)	−18.44 (0.0000)
EMD-EELM-EELM				2.46 (0.9930)

Table 6
DM test results for EEMD/EMD-EELM-ADD/EELM and EELM models in six-step-ahead prediction.

Tested model	Benchmark model			
	EEMD-EELM-EELM	EMD-EELM-ADD	EMD-EELM-EELM	EELM
EEMD-EELM-ADD	−13.11 (0.0000)	−6.11 (0.0000)	−15.27 (0.0000)	−16.00 (0.0000)
EEMD-EELM-EELM		5.45 (0.9999)	−5.29 (0.0000)	−6.51 (0.0000)
EMD-EELM-ADD			−11.24 (0.0000)	−12.35 (0.0000)
EMD-EELM-EELM				−1.97 (0.0246)

Table 7
DM test results for EEMD-EELM/ANN/LSSVR-ADD/EELM/ANN/LSSVR and EELM models in one-step-ahead prediction.

Tested model	Benchmark model					
	EEMD-ANN-ADD	EEMD-ANN-ANN	EEMD-LSSVR-ADD	EEMD-EELM-EELM	EEMD-LSSVR-LSSVR	EELM
EEMD-EELM-ADD	−2.87 (0.0020)	−23.96 (0.0000)	−8.70 (0.0000)	−19.43 (0.0000)	−15.28 (0.0000)	−24.59 (0.0000)
EEMD-ANN-ADD		−19.84 (0.0000)	−8.60 (0.0000)	−19.13 (0.0000)	−15.19 (0.0000)	−23.10 (0.0000)
EEMD-ANN-ANN			−7.63 (0.0000)	−17.32 (0.0000)	−14.66 (0.0000)	−21.61 (0.0000)
EEMD-LSSVR-ADD				−2.13 (0.0166)	−15.49 (0.0000)	3.98 (1.0000)
EEMD-EELM-EELM					−11.50 (0.0000)	10.35 (1.0000)
EEMD-LSSVR-LSSVR						12.77 (1.0000)

Table 8
DM test results for EEMD-EELM/ANN/LSSVR-ADD/EELM/ANN/LSSVR and EELM models in three-step-ahead prediction.

Tested model	Benchmark model					
	EEMD-ANN-ADD	EEMD-ANN-ANN	EEMD-LSSVR-ADD	EEMD-EELM-EELM	EEMD-LSSVR-LSSVR	EELM
EEMD-EELM-ADD	−3.81 (0.0001)	−13.57 (0.0000)	−8.44 (0.0000)	−13.87 (0.0000)	−11.28 (0.0000)	−21.37 (0.0000)
EEMD-ANN-ADD		−13.97 (0.0000)	−8.31 (0.0000)	−13.64 (0.0000)	−11.22 (0.0000)	−20.79 (0.0000)
EEMD-ANN-ANN			−7.84 (0.0000)	−12.61 (0.0000)	−11.02 (0.0000)	−20.09 (0.0000)
EEMD-LSSVR-ADD				2.01 (0.9780)	−9.62 (0.0000)	−0.70 (0.2416)
EEMD-EELM-EELM					−9.51 (0.0000)	−3.94 (0.0000)
EEMD-LSSVR-LSSVR						7.47 (1.0000)

Table 9
DM test results for EEMD-EELM/ANN/LSSVR-ADD/EELM/ANN/LSSVR and EELM models in six-step-ahead prediction.

Tested model	Benchmark model					
	EEMD-ANN-ADD	EEMD-ANN-ANN	EEMD-LSSVR-ADD	EEMD-EELM-EELM	EEMD-LSSVR-LSSVR	EELM
EEMD-EELM-ADD	−2.55 (0.0054)	−2.19 (0.0142)	−8.32 (0.0000)	−13.11 (0.0000)	−11.37 (0.0000)	−16.00 (0.0000)
EEMD-ANN-ADD		0.54 (0.7058)	−8.18 (0.0000)	−12.66 (0.0000)	−11.23 (0.0000)	−15.70 (0.0000)
EEMD-ANN-ANN			−8.23 (0.0000)	−12.79 (0.0000)	−11.26 (0.0000)	−15.76 (0.0000)
EEMD-LSSVR-ADD				0.50 (0.6929)	−9.65 (0.0000)	−4.84 (0.0000)
EEMD-EELM-EELM					−9.45 (0.0000)	−6.51 (0.0000)
EEMD-LSSVR-LSSVR						5.62 (1.0000)

the added step in EEMD, i.e., running 100 EMDs with different noise levels and then calculating the ensemble average.

From the results of DM test shown in Tables 4–9, five similar findings can be found. First, when the proposed EEMD-EELM-ADD model is treated as the testing target, the p-values are all far smaller than 10%, indicating that the proposed ensemble learning paradigm performs statistically better than all other models in all cases, under the confidence level of 90%. Second, according to Tables 4–6, EEMD-EELM-ADD and EEMD-EELM-EELM can yield better results than EMD-EELM-ADD and EMD-EELM-EELM at the significance level of 10%, verifying the superiority of the EEMD method to its original form EMD in data decomposition. Third, EEMD-EELM-EELM is obviously inferior to EEMD/EMD-EELM-ADD, indicating that the simple but efficient ADD approach is a quite effective ensemble tool in formulating decomposition and ensemble methodologies. Fourth, according to Tables 7–9, the EEMD-EELM-EELM can outperform the other ensemble learning paradigms without EELM, while EEMD-LSSVR-LSSVR/ADD performs the poorest at 90% confidence level in most cases. This indicates that the LSSVR without a careful parameters optimization might lose power in individual and ensemble prediction even under the framework of "decomposition and ensemble". However, such process of parameters selection for LSSVR tends to take a lot of time in large dataset case. Finally, EEMD/EMD-EELM-ADD can be proved to defeat the single EELM method under the confidence level of 90%, indicating the decomposition-and-ensemble framework can significantly improve the prediction function.

4.2.3. Robust analysis

Using random initial settings, the EELM, ELM and ANN methods generate different prediction results in different runs. Therefore, instability may be one important problem in these three neural network (NN) models, which should be considered in model

Table 10
Robustness analysis for single NN models.

Horizon	One			Three			Six		
Model	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*
EELM	**0.0039**	**0.0000**	**0.0071**	**0.0014**	**0.0000**	**0.0064**	**0.0011**	**0.0000**	**0.0037**
ELM	0.0121	0.0005	0.1088	0.0065	0.0016	0.3351	0.0027	0.0011	0.3139
ANN	0.0176	0.0029	0.3132	0.0087	0.0019	0.1905	0.0057	0.0011	0.1227

Table 11
Robustness analysis for ensemble learning models with different NNs.

Horizon	One			Three			Six		
Model	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*	*Std.* of D_{stat}	*Std.* of *MAPE*	*Std.* of *RMSE*
EEMD-EELM-ADD	**0.0066**	**0.0001**	**0.0094**	**0.0061**	**0.0001**	0.0234	0.0065	**0.0003**	**0.1431**
EEMD-EELM-EELM	0.0350	0.0064	0.5635	0.0555	0.0062	0.4945	0.0441	0.0050	0.3951
EEMD-ANN-ADD	0.0091	0.0002	0.0158	0.0079	0.0001	**0.0135**	0.0098	0.0003	0.0312
EEMD-ANN-ANN	0.0193	0.0009	0.0663	0.0143	0.0006	0.0498	**0.0055**	**0.0003**	0.0246

performance evaluation. Accordingly, we run all models of EELM, ELM, ANN, EEMD-EELM-ADD, EEMD-EELM-EELM, EEMD-ANN-ADD and EEMD-ANN-ANN for ten times, and analyze their robustness in terms of standard deviations (std.) of D_{stat}, *MAPE* and *RMSE*. Tables 10 and 11 report the results, and some similar conclusions can be summarized. First, focusing on single models, it is easy to find that EELM model is the most stable model, since its standard deviations of D_{stat}, *MAPE* and *RMSE* are far less than those of ELM and ANN methods. Second, in the case of ensemble learning models, the most robust model is EEMD-EELM-ADD learning paradigm, and the standard deviations of D_{stat}, *MAPE* and *RMSE* are far less than 0.01 in all cases, while EEMD-EELM-EELM is the poorest model. The main reason can be referred to the poor ensemble forecasting performance of EELM. The results further indicate that the proposed EEMD-EELM-ADD learning paradigm, with stable individual forecasting tool of EELM and simple but efficient ensemble tool of ADD, is a quite stable and robust prediction technique for crude oil price data.

4.2.4. Summarizations

From the above discussions, we can draw the following six main conclusions:

1) The proposed EEMD-based EELM ensemble learning methodology is significantly superior to all other methods listed in the study in terms of both level prediction and direction accuracy.
2) Amongst single models, EELM performs the best, indicating its effectiveness in crude oil price forecasting relative to other AI tools and traditional method.
3) The ensemble learning methods perform strikingly better than the single models in most cases, indicating that the strategy of "decomposition and ensemble" can effectively improve the prediction performance for crude oil price data.
4) The EEMD-based EELM ensemble learning paradigm is an effective and stable forecasting technique, in terms of time saving and robustness.
5) Due to nonlinearity and nonstationarity, nonlinear AI models with powerful computer learning are much more suitable for crude oil price prediction than linear traditional method.
6) Finally, the proposed EEMD-based EELM ensemble learning algorithm, with powerful strategy of "decomposition and ensemble" and competitive AI tool of EELM as individual forecasting model, can be used as a promising tool for crude oil price forecasting, in terms of prediction accuracy, time-saving and robustness.

5. Conclusions

Due to the intrinsic complexity of crude oil price data in terms of its interactive involving factors, a novel decomposition-and-ensemble learning paradigm, integrating ensemble empirical mode decomposition (EEMD) and extended extreme learning machine (EELM), is proposed for crude oil price forecasting, based on the principle of "decomposition and ensemble". The empirical study shows that the proposed ensemble learning paradigm can significantly improve prediction performance and statistically outperform some other popular forecasting methods (including single popular forecasting tools and similar ensemble learning paradigms) in terms of prediction accuracy, time saving and robustness. This further indicates that the proposed EEMD-based EELM ensemble learning paradigm with effective decomposition algorithm (i.e., EEMD) and powerful, fast and stable AI individual forecasting tool (i.e., EELM), can be used as a very promising method for complex time series forecasting problems, especially for crude oil price with high volatility and irregularity.

Besides crude oil price, the proposed EEMD-based EELM ensemble learning paradigm can be also used for other tough forecasting tasks. Furthermore, this study only considers univariate time series analysis, while some extra factors affecting crude oil market can be also taken into consideration to enhance the prediction capability of the novel model. We will look into these issues in the near future.

Acknowledgments

This work is partially supported by grants from the National Science Fund for Distinguished Young Scholars (NSFC no. 71025005), the National Natural Science Foundation of China (NSFC no. 91224001 and NSFC no. 71301006), the National Program for Support of Top-Notch Young Professionals and the Fundamental Research Funds for the Central Universities in BUCT.

References

Abdullah, S.N., Zeng, X., 2010. Machine learning approach for crude oil price prediction with Artificial Neural Networks-Quantitative (ANN-Q) model. In:

Proceedings of IEEE the 2010 International Joint Conference on Neural Networks, pp. 1–8.
Bao, Y., Zhang, X., Yu, L., Lai, K.K., Wang, S., 2007. Hybridizing wavelet and least squares support vector machines for crude oil price forecasting. In: Proceedings of the 2nd International Workshop on Intelligent Finance, pp. 1–15.
Box, G.E.P., Jenkins, G., 1970. Time Series Analysis: Forecasting and Control, 1st ed. Holden Day Press, San Francisco.
Chen, S.S., 2009. Revisiting the inflationary effects of oil prices. Energy J., 141–154.
Chen, S.S., Hsu, K.W., 2012. Reverse globalization: does high oil price volatility discourage international trade? Energy Econ. 34 (5), 1634–1643.
Deng, J., 1982. Control problems of grey systems. Systems & Control Letters 1 (5), 288–294.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.
Gholamian, M.R., Ghomi, S.F., Ghazanfari, M., 2005. A hybrid systematic design for multiobjective market problems: a case study in crude oil markets. Eng. Appl. Artif. Intell. 18 (4), 495–509.
He, K., Yu, L., Lai, K.K., 2012. Crude oil price analysis and forecasting using wavelet decomposed ensemble model. Energy 46 (1), 564–574.
Hou, A., Suardi, S., 2012. A nonparametric GARCH model of crude oil price return volatility. Energy Econ. 34 (2), 618–626.
Huang, G.B., Chen, L., Siew, C.K., 2006a. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Netw. 17 (4), 879–892.
Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In: Proceedings of 2004 IEEE International Joint Conference on Neural Networks, vol. 2, pp. 985–990.
Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006b. Extreme learning machine: theory and applications. Neurocomputing 70 (1), 489–501.
Huang, N.E., Shen, Z., Long, S.R., et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc. R. Soc. Lond. Ser. A: Math., Phys. Eng. Sci. 454, 903–995.
Huntington, H.G., 1994. Oil price forecasting in the 1980s: what went wrong? Energy J., 1–22.
Kaboudan, M.A., 2001. Compumetric forecasting of crude oil prices. In: Proceedings of IEEE the 2001 Congress on Evolutionary Computation, vol. 1, pp. 283–287.
Khashman, A., Nwulu, N.I., 2011. Intelligent prediction of crude oil price using Support Vector Machines. In: Proceedings of 2011 IEEE 9th International Symposium on Applied Machine Intelligence and Informatics, pp. 165–169.
Kulkarni, S., Haidar, I., 2009. Forecasting model for crude oil price using artificial neural networks and commodity futures prices. arXiv preprint arXiv: 0906.4838.
Lanza, A., Manera, M., Giovannini, M., 2005. Modeling and forecasting cointegrated relationships among heavy oil and product prices. Energy Econ. 27 (6), 831–848.
Li, S., Ge, Y., 2013. Crude oil price prediction based on a dynamic correcting support vector regression machine. Abstract and Applied Analysis 2013, 1–7.
Li, Z., Sun, J., Wang, S., 2013. An information diffusion-based model of oil futures price. Energy Econ. 36, 518–525.
Lin, A., 2009. Prediction of international crude oil futures price based on GM (1,1). In: Proceedings of IEEE International Conference on Grey Systems and Intelligent Services, pp. 692–696.
Liu, H., Tian, H.Q., Pan, D.F., Li, Y.F., 2013. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. Appl. Energy 107, 191–208.
Liu, X., Gao, C., Li, P., 2012. A comparative analysis of support vector machines and extreme learning machines. Neural Netw. 33, 58–66.
Lu, C.J., Shao, Y.E., 2012. Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. Math. Probl. Eng. 2012, 1–15.
Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., Lendasse, A., 2010. OP-ELM: optimally pruned extreme learning machine. IEEE Trans. Neural Netw. 21 (1), 158–162.
Mohammadi, H., Su, L., 2010. International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. Energy Econ. 32 (5), 1001–1008.
Murat, A., Tokat, E., 2009. Forecasting oil price movements with crack spread futures. Energy Econ. 31 (1), 85–90.
Oman, H., 2003. Energy sources for the world's post-petroleum era. IEEE Aerosp. Electron. Syst. Mag. 18 (11), 35–39.
Pati, J.C., Mishra, K.K., Patnaik, S.K., 2013. A comparative study on short term load forecasting using BPNN and extreme learning machine. Int. J. Adv. Res. Sci. Technol. 2 (1), 30–34.
Quan, T., Liu, X., Liu, Q., 2010. Weighted least squares support vector machine local region method for nonlinear time series prediction. Appl. Soft Comput. 10 (2), 562–566.
Rao, C.R., Mitra, S.K., 1971. Generalized Inverse of Matrices and its Applications, seventh ed. Wiley, New York.
Rong, H.J., Ong, Y.S., Tan, A.H., Zhu, Z., 2008. A fast pruned-extreme learning machine for classification problem. Neurocomputing 72 (1), 359–366.
Shrivastava, N.A., Panigrahi, B.K., 2014. A hybrid wavelet-ELM based short term price forecasting for electricity markets. Int. J. Electr. Power Energy Syst. 55, 41–50.
Singh, R., Balasundaram, S., 2007. Application of extreme learning machine method for time series analysis. Int. J. Intell. Technol. 2 (4), 256–262.
Sun, Z.L., Choi, T.M., Au, K.F., Yu, Y., 2008. Sales forecasting using extreme learning machine with applications in fashion retailing. Decis. Support Syst. 46 (1), 411–419.
Sun, Z., Au, K.F., Choi, T.M., 2007. A neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines. IEEE Trans. Syst. Man, Cybern., Part B: Cybern. 37 (5), 1321–1331.
Tang, L., Dai, W., Yu, L., Wang, S.Y., 2015. A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. Int. J. Inf. Technol. Decis. Mak. 14 (1), 141–169.
Tang, L., Wang, S., Yu, L., 2011. EEMD-LSSVR-based decomposition-and-ensemble methodology with application to nuclear energy demand forecasting. In: Proceedings of IEEE 2011 Fourth International Joint Conference on Computational Sciences and Optimization, pp. 589–593.
Tang, L., Yu, L., He, K.J., 2014. A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. Appl. Energy 128, 1–14.
Tang, L., Yu, L., Wang, S., Li, J., Wang, S., 2012. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. Appl. Energy 93, 432–443.
Tian, H., Meng, B., 2010. A new modeling method based on bagging ELM for day-ahead electricity price prediction. In: Proceedings of 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications, pp. 1076–1079.
Wang, J., Zhang, W., Li, Y., Wang, J., Dang, Z., 2014. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl. Soft Comput. 23, 452–459.
Wang, S.Y., Yu, L., Lai, K.K., 2005. Crude oil price forecasting with TEI@I methodology. J. Syst. Sci. Complex. 18 (2), 145–166.
Wang, S., Yu, L., Tang, L., Wang, S., 2011. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. Energy 36 (11), 6542–6554.
Wang, X., Han, M., 2015. Improved extreme learning machine for multivariate time series online sequential prediction. Eng. Appl. Artif. Intell. 40, 28–36.
Wong, W.K., Guo, Z.X., 2010. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. Int. J. Prod. Econ. 128 (2), 614–624.
Wu, C.L., Chau, K.W., 2010. Data-driven models for monthly streamflow time series prediction. Eng. Appl. Artif. Intell. 23 (8), 1350–1367.
Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 1 (1), 1–41.
Xie, G., Wang, S., Zhao, Y., Lai, K.K., 2013. Hybrid approaches based on LSSVR model for container throughput forecasting: a comparative study. Appl. Soft Comput. 13 (5), 2232–2241.
Xie, Q., Xuan, B., Peng, S., Li, J., Xu, W., Han, H., 2008. Bandwidth empirical mode decomposition and its application. Int. J. Wavelets, Multiresolution Inf. Process. 6 (6), 777–798.
Xie, W., Yu, L., Xu, S., Wang, S., 2006. A new method for crude oil price forecasting based on support vector machines. In: Alexandrov, V.N., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (Eds.), Computational Science – ICCS 2006. Springer, Berlin, Heidelberg, pp. 444–451.
Yu, L., Lai, K.K., Wang, S., He, K., 2007. Oil price forecasting with an EMD-based multiscale neural network learning paradigm. In: Shi, Y., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (Eds.), Computational Science – ICCS 2007, 2007. Springer, Berlin, Heidelberg, pp. 925–932.
Yu, L., Wang, S., Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Econ. 30 (5), 2623–2635.
Yu, L., Wang, S., Lai, K.K., 2009. A neural-network-based nonlinear metamodeling approach to financial time series forecasting. Appl. Soft Comput. 9 (2), 563–574.
Yu, L., Zhao, Y., Tang, L., 2014. A compressed sensing based AI learning paradigm for crude oil price forecasting. Energy Econ. 46, 236–245.
Zhang, J., Yan, R., Gao, R.X., Feng, Z., 2010. Performance enhancement of ensemble empirical mode decomposition. Mech. Syst. Signal Process. 24 (7), 2104–2123.
Zhang, X., Lai, K.K., Wang, S.Y., 2008. A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Econ. 30 (3), 905–918.
Zhang, X., Yu, L., Wang, S., Lai, K.K., 2009. Estimating the impact of extreme events on crude oil price: an EMD-based event analysis method. Energy Econ. 31 (5), 768–778.