Applied Energy 220 (2018) 480–495

Contents lists available at ScienceDirect

Applied Energy

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

A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms

Minggang Wang[a,b,c], Longfeng Zhao[c], Ruijin Du[c,d], Chao Wang[c,e], Lin Chen[c,f], Lixin Tian[a,d,*], H. Eugene Stanley[c]

[a] School of Mathematical Science, Nanjing Normal University, Nanjing 210042, Jiangsu, China
[b] Department of Mathematics, Nanjing Normal University Taizhou College, Taizhou 225300, Jiangsu, China
[c] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
[d] Energy Development and Environmental Protection Strategy Research Center, Jiangsu University, Zhenjiang 212013, Jiangsu, China
[e] College of Economics and Management, Beijing University of Technology, Beijing 100124, China
[f] School of Management, Northwestern Polytechnical University, Xi'an 710072, Shanxi, China

HIGHLIGHTS

- A novel prediction paradigm (DFN-AI) is proposed based on complex network and AI algorithms.
- DFN analysis technique is performed to extract the fluctuation features in original data.
- A new data reconstruction method is designed by using the extracted data.
- A certain artificial intelligence tool is employed to model the reconstructed data.
- Empirical results demonstrate the effectiveness and robustness of DFN-AI method.

ARTICLE INFO

Keywords:
Complex network
Artificial intelligence algorithms
Crude oil price prediction

ABSTRACT

Forecasting the price of crude oil is a challenging task. To improve this forecasting, this paper proposes a novel hybrid method that uses an integrated data fluctuation network (DFN) and several artificial intelligence (AI) algorithms, named DFN-AI model. In the proposed DFN-AI model, a complex network time series analysis technique is performed as a preprocessor for the original data to extract the fluctuation features and reconstruct the original data, and then an artificial intelligence tool, e.g., BPNN, RBFNN or ELM, is employed to model the reconstructed data and predict the future data. To verify these results we examine the daily, weekly, and monthly price data from the crude oil trading hub in Cushing, Oklahoma. Empirical results demonstrate that the proposed DFN-AI models (i.e., DFN-BP, DFN-RBF, and DFN-ELM) perform significantly better than their corresponding single AI models in both the direction and level of prediction. This confirms the effectiveness of our proposed modeling of the nonlinear patterns hidden in crude oil prices. In addition, our proposed DFN-AI methods are robust and reliable and are unaffected by random sample selection, sample frequency, or breaks in sample structure.

1. Introduction

Because crude oil is a basic energy source and its price volatilities strongly impact a country's economic development, social stability, and national security [1], accurately predicting crude oil price fluctuations is a consistently active topic of research. The research on crude oil price fluctuations being carried out internationally is made more complex by the interplay among many factors—including market supply and demand [2], the US dollar exchange rate [3], speculative trading [4], geopolitical conflicts [5], and natural disasters [6]—that introduces a high level of noise into the crude oil data. Thus the crude oil prices, which exhibit such complex volatility characteristics as nonlinearity and uncertainty, are difficult to forecast and any results obtained uncertain. Therefore, crude price prediction remains a huge challenge.

Up to now, there has been a raft of literature discussing crude oil price forecasting. Among these prediction models, one of the most

* Corresponding author at: School of Mathematical Science, Nanjing Normal University, Nanjing 210042, Jiangsu, China.
E-mail addresses: magic821204@sina.com (M. Wang), zlfccnu@mails.edu.cn (L. Zhao), dudo999@126.com (R. Du), chaowanghn@vip.163.com (C. Wang), clnwpu@126.com (L. Chen), tianlx@ujs.edu.cn (L. Tian), hes@bu.edu (H. Eugene Stanley).

https://doi.org/10.1016/j.apenergy.2018.03.148
Received 4 January 2018; Received in revised form 5 March 2018; Accepted 28 March 2018

Nomenclature

X	original time series
N	data size
P	fluctuation series of X
S	symbol series
k	number of the symbols
s_i	symbol
L	length of the sliding window
l	sliding step
r	threshold
FM_i	the ith fluctuation modes
$\widehat{M}$	number of the fluctuation modes
M	number of different fluctuation modes
v_i^t	node numbered i at time t
$V_{i\to j}^{t+1}$	set of all out-neighbor nodes of v_i^t
W	weight
η	learning rate
E	the gradient of error function
B_i	the prototype of the input vectors
σ_i	the width of RBF unit i
$\widehat{X}$	predicted data
$f(x)$	the activation function
b_i	the bias of hidden node i
β_i	the weights of hidden neuron i to output neurons
EX	extracted data
SX	sub data of original data
α	the selectivity coefficient
RX	the reconstructed data

Abbreviations

DFN	data fluctuation network
BPNN	back propagation neural network
RBFNN	radial basis function neural network
ELM	extreme learning machine
DFN-BP	hybrid model based on DFN and BPNN
DFN-RBF	hybrid model based on DFN and RBFNN
DFN-ELM	hybrid model based on DFN and ELM
MAPE	mean absolute percentage error
RMSE	root mean square error
Dstat	directional statistic
DMS	Diebold-Mariano statistic

important models is econometric model. For instance, Lanza et al. [7] used cointegration and error correction models (ECM) to predict crude oil prices from January 2002 to June 2002. Murat et al. [8] proposed a vector error correction model (VECM) to forecast oil price movements and crack spread futures. Baumeister et al. [9] used vector autoregressive (VAR) to forecast WTI spot price. Xiang et al. [10] used an autoregressive integrated moving average (ARIMA) model to predict the Brent crude oil. Sadorsky [11] used several GARCH models to forecast the daily volatility in petroleum futures price returns. Fan et al. [12] introduced GARCH type models based on generalized error distribution (GED) to examine the risk spillover effect between West Texas Intermediate (WTI) and Brent crude oil markets. Kang et al. [13] then proposed a variety of conditional volatility models, including GARCH, IGARCH, CGARCH, and FIGARCH, to forecast the volatility of crude oil markets, and found that the CGARCH and FIGARCH models can forecast volatility persistence. Mohammadi et al. [14] investigated the out-of-sample forecasting performance of four volatility models—GARCH, EGARCH, APARCH and FIGARCH over January 2009 to October 2009. Hou and Suardi [15] focused on two crude oil markets, Brent and WTI, considered an alternative approach involving nonparametric method to model and forecast oil price return volatility. The main results of the above mentioned econometric models are listed in Table 1 (the upper part). In essence there are two different types of econometric models. The first is a structural model of the price of oil, including ECM [7], VECM [8], VAR [9] et al., depending on fundamental data such as demand and supply and is implemented through the use of a linear regression. This structural modeling approach includes explanatory variables other than just the past data of oil prices into the process. The second is a time series approach, including ARIMA [10], GARCH-type models [11–15] et al., only looking at the history of price to determine future price movement. Because they are able to capture time-varying volatility, econometric models have improved the accuracy of forecasting, but because they assume the data to be stationary, regular, and linear they cannot accurately model time series that are complex, irregular, and nonlinear [7–15].

In addition to the classic econometric approaches, artificial intelligence (AI) methods have been used to uncover the inner complexity of oil prices. For example, Moshiri et al. [16] set up a nonlinear and flexible artificial neural network (ANN) model to forecast daily crude oil futures prices traded at the New York Mercantile Exchange (NYMEX). Kaboudan [17] evaluated forecasts produced by two competing compumetric forecasting methods: genetic programming (GP) and artificial neural networks (ANN). Mostafa et al. [18] forecasted oil prices using gene expression programming (GEP) and artificial neural network (ANN) models. Kaboli et al. [19,20] developed artificial cooperative search algorithm (ACSA) and GEP to provide better-fit solution and improve the accuracy of estimation. Xie et al. [21] proposed a support vector machine (SVM) to forecast crude oil prices and compared its performance with ARIMA and back propagation neural network (BPNN). Shin et al. [22] employed semi-supervised learning (SSL) to forecast the upward and downward movement of oil prices. Yusof et al. [23] proposed least squares support vector machine (LSSVM) method of the oil futures price forecasting. Zhao et al. [24] introduced deep learning approach (SDAE) for WTI crude oil spot price forecasting. The main results of the above mentioned AI models are listed in Table 1 (the middle part). Unlike econometric models [7–15], artificial intelligence methods are able to model such complex characteristics as nonlinearity and volatility. Artificial intelligence methods also have disadvantages, For example, ANN and BPNN often suffer from local minima and over-fitting, while other AI models, such as SVM and GP including ANN, are sensitive to parameter selection [16–24].

Because single prediction models—including both econometric models and AI methods—are limited, many studies are now using hybrid methods to forecast crude oil prices. Some typical literature regarding the hybrid methods for crude oil price forecasting can be found in Table 1 (the bottom part). Overall, the hybrid methods often imply the combination of interdisciplinary methods to use their strengths and can be roughly classified into two categories: (1) the combination among AI models, such as the empirical mode decomposition (EMD) based neural network ensemble learning paradigm [25], the hybrid model combining the dynamic properties of multilayer back propagation neural network and the recent Harr A trous wavelet decomposition, i.e., HTW-MPNN [26], the hybrid model built upon EMD based on the feed-forward neural network (FNN) modeling framework incorporating the slope based method (SBM), i.e., EMD-SBM-FNN [27], a decomposition-and-ensemble learning paradigm integrating ensemble empirical mode decomposition (EEMD) and extended extreme learning machine (EELM), i.e., EEMD-EELM [28], the compressed sensing based learning paradigm, integrating compressed sensing based de-noising (CSD) and certain artificial intelligence (AI), i.e., CSD-AI [29], the alternative approach based on a genetic algorithm and neural network (GA-NN) [30], the hybrid AI system framework integrating web-based