Energy Economics 66 (2017) 9–16

Contents lists available at ScienceDirect

Energy Economics

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

A deep learning ensemble approach for crude oil price forecasting

Yang Zhao [a,b], Jianping Li [a,b,*], Lean Yu [c]

[a] Institute of Policy and Management, Chinese Academy of Sciences, Beijing, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] School of Economics and Management, Beijing University of Chemical Technology, Beijing, China

ARTICLE INFO

Article history:
Received 7 December 2016
Received in revised form 27 March 2017
Accepted 29 May 2017
Available online 9 June 2017

JEL classification:
C45
C52
C53
Q47
E37
C63

Keywords:
Crude oil price forecasting
Deep learning
Ensemble learning
Stacked denoising autoencoder
Bagging
Multivariate forecasting

ABSTRACT

As crude oil price is influenced by numerous factors, capturing its behavior precisely is quite challenging, and thus leads to the difficulty of forecasting. In this study, a deep learning ensemble approach is proposed to deal with this problem. In our approach, two techniques are utilized. One is an advanced deep neural network model named stacked denoising autoencoders (SDAE) which is used to model the nonlinear and complex relationships of oil price with its factors. The other is a powerful ensemble method named bootstrap aggregation (bagging) which generates multiple data sets for training a set of base models (SDAEs). Our approach combines the merits of these two techniques and is especially suitable for oil price forecasting. In the empirical study, the WTI crude oil price series are investigated and 198 economic series are used as exogenous variables. Our approach is tested against some competing approaches and shows superior forecasting ability that is statistically proved by three tests.

1. Introduction

Crude oil price volatilities have great impact on the economic activities of the world from many aspects. It has been largely free to fluctuate in response to the forces of supply and demand (Baumeister and Kilian, 2016). Besides these two fundamentals, various factors strike the oil prices at different frequencies. In energy market, production of other commodities including natural gas, coal and renewable energy, may have substitution effect which leads to the volatility of oil price indirectly. Other factors such as financial markets, economic growth, technology development and irregular events also influence the oil price in different ways. Complex relationships are built between these factors and oil prices, thus drive strong fluctuations in crude oil market. As a result, forecasting oil price has always been a tough task. However, seeking for promising forecasting approaches for oil price series is hardly outdated since crude oil is the main source of energy in the world and dominates the economic activities. Accurate forecast of oil price guides the decision making of many sectors such as business organizations and governments.

Research on crude oil price forecasting has lasted for decades and plenteous approaches have been proposed. Besides the classic econometric approaches, various machine learning methods are utilized to mining the inner complexity of oil price. The most typical and commonly used machine learning methods include neural network (NN) and support vector machine (SVM). They are particularly welcomed for their capability of modeling complex characteristics such as nonlinearity and volatility. For example, genetic algorithm (GA) (Kaboudan, 2001), NN (Moshiri and Foroutan, 2006) and SVM (Xie et al., 2006) are first applied to forecast crude oil price in earlier studies. In recent years, semi-supervised learning (SSL) (Shin et al., 2013), gene expression programming (GEP) (Mostafa and El-Masry, 2016) are used for oil price forecasting. The above mentioned models are single models of original form while hybridization of single machine learning models (especially NN) for oil price forecasting is becoming a popular phenomenon. Hybrid models have proved to have better forecasting accuracies than their corresponding single machine learning techniques. For example, Godarzi et al. (2014) forecast oil price with an NN based dynamic

* Corresponding author at: Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: ljp@casipm.ac.cn (J. Li).

http://dx.doi.org/10.1016/j.eneco.2017.05.023
0140-9883/

nonlinear autoregressive with exogenous input (NARX) model which is a multivariate forecasting model. Yu et al. (2014) propose a hybrid forecasting model in which the training data is first preprocessed by compressed sensing denoising and then it is used for training certain machine learning techniques including NN and support vector regression (SVR). Ghaffari and Zare (2009) propose an adaptive network-based fuzzy inference system (ANFIS) model for forecasting daily oil price which is first preprocessed with a data filtering algorithm. Chiroma et al. (2015) proposes a hybrid crude oil price forecasting model in which the mataparameters of NN are selected by GA. There is a special kind of hybrid models that is actually ensemble learning paradigm in which the prediction is the integration of the output of several heterogeneous or homogenous models. For instance, Gabralla et al. (2013) proposes an ensemble oil price forecasting model based on SVR, instance based learning (IBL) and K star, and the prediction is generated by taking an average of all the individual forecasts of these machine learning techniques. There are also ensemble models that first decomposed oil price series into several components and then combine the forecasts of each components generated by NNs (Jammazi and Aloui, 2012; Xiong et al., 2013; Yu et al., 2008b, 2016). Essentially, a hybrid learning paradigm consists of two parts. One is a core machine learning technique that is used for training and forecasting the oil price. The other is an additional technique that is used for enhancing the forecasting ability of the entire model (i.e., improve the generalization ability of the core machine learning technique or break the forecasting task into simpler ones).

So far, the machine learning techniques in the above mentioned forecasting models are shallow architectures (e.g. NN with only one hidden layer). Bengio (2009) points out that the functions cannot be efficiently represented (in terms of number of tunable elements) by architectures that are too shallow. In terms of oil price forecasting, shallow architecture base forecasting approaches may fail to model the complex patterns and volatile behaviors of oil price driven by numerous factors. Therefore, a natural idea is to model the oil price with deep architecture based approaches. A deep architecture is the composition of multiple levels of non-linear operations. Recently, deep learning (DL) is becoming a mainstream of machine learning technique, and it has shown strong capacity in various nonlinear modeling tasks such as classification and feature extraction. Hinton and Salakhutdinov (2006) propose a greedy layer wise training strategy which solves the training problem in deep neural network (DNN). After that, relevant algorithms and applications are becoming flourishing. It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government (LeCun et al., 2015). It is worth mentioning that, to the best of our knowledge, no DL approach has ever been applied in oil price forecasting. Thus in this work, a deep learning ensemble (DLE) approach for forecasting oil price is proposed by applying one of the most popular DNN model named stacked denoising autoencoders (SDAE) (Vincent et al., 2010) as base model. In the DLE approach, the ensemble method named bootstrap aggregation (bagging) (Breiman, 1996a) is used to generate multiple data set for training SDAEs. Then the trained SDAEs are used to generate predictions. Finally, the predictions of base models are averaged to form the final prediction. For multivariate forecasting, a large set of economic series (e.g., oil production and consumption, stock levels, macroeconomic and financial indicators) are considered as exogenous variables. All these series have ever been used in previous studies (Godarzi et al., 2014; Naser, 2016; Ye et al., 2006; Zagaglia, 2010). In general, a multivariate deep learning ensemble approach is proposed for crude oil price forecasting. Two main contributions of this paper are presented as follows. (1) For the first time, a deep learning approach is introduced to mine the complex relationship of oil price with exogenous variables. (2) A novel deep learning ensemble forecasting approach is built for forecasting oil price series.

The remaining part of this paper is organized as follows. Section 2 describes the formulation of our approach. Section 3 reports the experimental results. Section 4 concludes the paper and outlines the future research direction.

2. Methodology formulation

In this section, a novel deep learning ensemble approach named SDAE bagging (SDAE-B) is formulated for crude oil price forecasting. Deep learning approach (SDAE) and ensemble learning approach (bagging) are respectively introduced in Sections 2.1 and 2.2. Multivariate forecasting approach is described in Section 2.3. Finally, the overall process of the proposed approach is presented in Section 2.4.

2.1. Stacked denoising autoencoders

SDAE (Vincent et al., 2008, 2010) is a popular DNN model that has been proved to have higher prediction accuracy than some competing machine learning models such as SVM, stacked autoencoders (SAE) and deep belief networks (DBN) in a range of classification problems. SDAE is built by stacking several denoising autoencoders (DAEs) which is a special kind of neural network structure. To illustrate the SDAE, autoencoder (AE) and DAE are first introduced.

AE is a one hidden layer neural network where its input and output size are equal. It first maps an input vector $\mathbf{x} \in [0,1]^d$ to a hidden representation $\mathbf{y} \in [0,1]^{d'}$ through a deterministic function:

$$\mathbf{y} = f_\theta(\mathbf{x}) = \phi_f(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{1}$$

where $f(\mathbf{x})$ is parameterized by $\theta = \{\mathbf{W}, \mathbf{b}\}$, $\mathbf{W}$ is a $d' \times d$ weight matrix, $\mathbf{b}$ is a bias vector, and $\phi_f(\cdot)$ is a nonlinear activation function. Then, the representation $\mathbf{y}$ is mapped back to vector $\mathbf{z} \in [0,1]^d$ in input space:

$$\mathbf{z} = g_{\theta'}(\mathbf{y}) = \phi_g(\mathbf{W}'\mathbf{y} + \mathbf{b}') \tag{2}$$

with $\theta' = \{\mathbf{W}', \mathbf{b}'\}$. Each training $\mathbf{x}^{(i)}$ is thus mapped to a corresponding $\mathbf{y}^{(i)}$ and a reconstruction $\mathbf{z}^{(i)}$. The parameter of this model is optimized to minimize the average reconstruction error.

$$\theta^*, \theta'^* = \arg\min_{\theta,\theta'} \frac{1}{n}\sum_{i=1}^{n} L\left(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}\right) = \arg\min_{\theta,\theta'} \frac{1}{n}\sum_{i=1}^{n} L\left(\mathbf{x}^{(i)}, g_\theta\left(f_\theta\left(\mathbf{x}^{(i)}\right)\right)\right) \tag{3}$$

where L is loss function that can be either traditional squared error $L(\mathbf{x},\mathbf{z}) = \|\mathbf{x}-\mathbf{z}\|^2$ or reconstruction cross-entropy $L_H(x,z) = H(\mathbf{B}_\mathbf{x}\|\mathbf{B}_\mathbf{z}) = -\sum_{k=1}^{d}[\mathbf{x}_k \log \mathbf{z}_k + (1-\mathbf{x}_k)\log(1-\mathbf{z}_k)]$.

Vincent et al. (2010) point out that training a common autoencoder (AE) is unable to guarantee the extraction of useful features just by minimizing the loss function, while DAE can change the reconstruction criteria by denoising (i.e., cleaning partially corrupted input through AE) so as to learn a good representation. The core idea of DAE is to reconstruct a clean input from a corrupted version. Firstly, corrupt the initial input $\mathbf{x}$ into $\tilde{\mathbf{x}}$ by stochastic mapping $\tilde{x} \sim q_D(\tilde{\mathbf{x}}|\mathbf{x})$. Then the corrupted input $\tilde{\mathbf{x}}$ is mapped to a hidden representation $\mathbf{y} = f_\theta(\tilde{\mathbf{x}}) = \phi(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b})$. Finally, $\mathbf{y}$ is reconstructed to $\mathbf{z} = g_{\theta'}(\mathbf{y})$. For a training set, the best parameters θ and θ' are trained by minimizing the average reconstruction error between $\mathbf{z}$ and the uncorrupted input $\mathbf{x}$. The procedure is shown in Fig. 1.

The deep neural networks SDAE is built by stacking several DAEs, in the same way as stacking Restricted Boltzmann Machines (RBMs) in deep belief networks (Hinton et al., 2006; Hinton and Salakhutdinov, 2006). The procedure is depicted in Fig. 2. It follows the greedy layer wise pretraining procedure (Hinton and Salakhutdinov, 2006) of learning one layer of features at a time. Specifically, the first DAE is trained independently with the training set as its input and the mapping function $f_\theta^{(1)}$ is thus learnt. Then the second DAE is trained with the hidden representation y of the first DAE as its input and the mapping function f_θ