

Predicting Model of Crude Oil Price Combining Stochastic Time Effective and Factorization Machine Based Neural Network



Fang Wang and Menggang Li

Abstract This paper concentrates on the model, factorization machine based neural network (NN) with stochastic time effective function (FM-ST-NN), to predict the WTI crude oil price trend. The factorization machine technology makes the proposed model capable of grasping the interactive factorization among inputs. And the stochastic time effective function shows the effect of historical data's time on weights which reflects the impact on the predictive model. The forecasting results are compared with other methods including neural network and factorization machine based neural network. And the predicting performance are evaluated through the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) accuracy measures.

Keywords Factorization machine · Neural network · Crude oil price prediction · Time random effects

1 Introduction

Crude oil forecasting is always a significant topic which has drawn a large amount of attention from the area of economy and finance. As some difficulties for disobeying the statistical assumptions in dealing with nonlinear and non-stationary time series when considering the crude oil price accurately, the predicting becomes more challenging.

F. Wang (✉)

School of Economics and Management, Beijing Laboratory of National Economic Security
Early-warning Engineering, Beijing Jiaotong University, Beijing, China
e-mail: 98930192@bjtu.edu.cn

M. Li

National Academy of Economic Security, Beijing Laboratory of National Economic Security
Early-warning Engineering, Beijing Jiaotong University, Beijing, China
e-mail: mglil@bjtu.edu.cn

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2020

M. Li et al. (eds.), *IEIS2019*,

https://doi.org/10.1007/978-981-15-5660-9_18

There are numerous and effective models proposed for predicting crude oil price series. Besides some classical statistical models [1, 2], some models are developed based on machine learning technique in recent years [3–12]. From the machine learning approaches, support vector machine (SVM) [3–5] is one of the artificial intelligent based approaches which can capture the crude oil prices' volatility. Another popularly and frequently used model is neural network [6, 7]. Neural network has the advantages on capturing the nonlinear functional relation through the interconnected neurons and training modes. Some applicable types of neural networks like radial basis function neural network, recurrent neural network, wavelet neural network, etc [8–12] have also been utilized to predict financial time series. Consequently, hybrid methods [13, 14] and some neural network models combined with random jump or with random time effective [15–18] have been widely employed in forecasting.

Although the artificial intelligent based technologies have achieved markable results, there are still limitations. As far as we know, a large number of models do not pay attention to the interactions among the inputs with different scales of features. For example, most of the neural network models handle the non-linearities by the activation functions and they were widely used in some applications such as image processing, mechanical translation, speech recognition and so on. Factorization machine (FM) is firstly introduced by Rendle [19] who used it for collaborative recommendations. FM is an advantageous method in learning feature interactions. In contrast to the matrix factorization method which only model the relation between two entities [20], FM shows its flexibility. FM also has applications in industry and gives a promising direction for the prediction purpose in regression, classification and ranking [21–23].

In this paper, we combine factorization machine with neural network technique (FM-NN) for crude oil price forecasting. In order to show the effects of stochastic time on data in different time, we incorporate the stochastic time effective function [17] to modify the weights when computing the global error. It helps to modify the weights of the proposed neural network (FM-ST-NN) more effectively. To evaluate the model's performance, we select WTI crude oil data from January 3, 2012 to December 31, 2018. To show the advantages of the proposed FM-ST-NN model, we compare the predicting results with two neural network models including NN and FM-NN. The performance of different models are evaluated through three accuracy measures finally.

The rest of this paper is presented as follows. In Sect. 2, we give the prediction model FM-ST-NN. In this section, we first construct the model and introduce some needed ingredients. Then the algorithm of FM-ST-NN is given. Section 3 presents the main predicting results of FM-ST-NN model. The forecasting results shows more accuracy compared with the other two neural network models, i.e., NN and FM-NN. In Sect. 4, we highlight some necessary conclusions finally and give some directions for future work.

2 Methodology

2.1 FM-ST-NN (Factorization Machine-Neural Network with a Stochastic Time Effective Function)

In this section, we first describe a three layer neural network incorporating factorization machine ideas which is shown in Fig. 1. Compared with neural network, the main differences are located in the hidden layer from the structure. Be more concrete, the blue nodes receive data after the linear connections from the set of input neurons. And then compute the results through the nonlinear activation function' operation. But the red nodes compute the results through the activation function of the factorized interactions among the inputs.

The output of FM models all interactions between each pair of features, i.e., a linear interactions and an inner product of each feature. The estimating results of FM is

$$\begin{aligned} \hat{y}(x) &= w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \\ &= w_0 + \sum_{i=1}^n w_i x_i + \frac{1}{2} \sum_{f=1}^{k'} \left(\left(\sum_{i=1}^n v_{if} x_i \right)^2 - \sum_{i=1}^n v_{if}^2 x_i^2 \right), \end{aligned} \tag{1}$$

where k' is a user-specified dimension parameter for feature x_i .

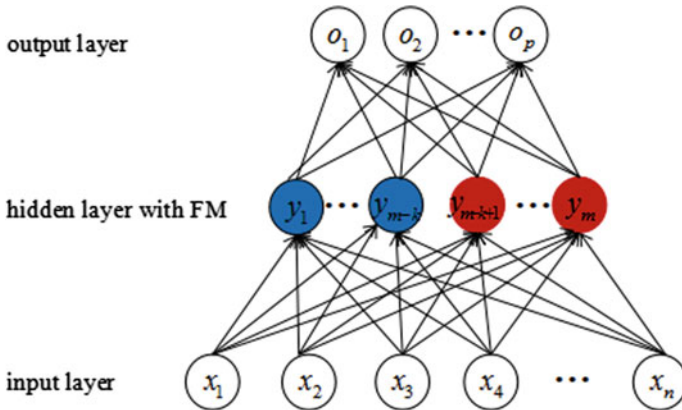


Fig. 1 The architecture of factorization machine-neural network

Suppose that there are n input features in the input layer, m neurons in the hidden layer and p output neurons in the output layer. Then, we describe the forecasting process of FM-ST-NN step by step below.

1). The hidden layer: the neurons in the hidden layer are partitioned into two parts:

$$y_j = f\left(\sum_{i=1}^n v_{ij}x_i\right), i = 1, 2, \dots, n, j = 1, 2, \dots, m - k, \quad (2)$$

$$y_{m-k+j} = f\left(\frac{1}{2}\left(\left(\sum_{i=1}^n \hat{v}_{i,m-k+j}x_i\right)^2 - \sum_{i=1}^n (\hat{v}_{i,m-k+j})^2x_i^2\right)\right), \\ i = 1, 2, \dots, n, j = 1, 2, \dots, k, \quad (3)$$

where x_i is the value of i th node, y_j denotes the value of the j th node, v_{ij} is the linear weight relates the i th input node to the j th output node which is undetermined, \hat{v}_{ij} is the latent weight corresponding to the i th feature which is also undetermined, k is the user-specified dimension, and f is the activation function.

2). The output layer:

$$O_j = f\left(\sum_{i=1}^m w_{ij}y_j\right), j = 1, 2, \dots, p, \quad (4)$$

where w_{ij} is the connection weight which is undetermined, O_j is the value of the j th node in the output layer.

3). The loss layer: the squared loss function is a commonly used loss function and the error of the sample e_{i_0} is computed as

$$e_{i_0}(t) = \frac{1}{2}\phi(t_{i_0})\sum_{j=1}^p (T_j(t) - O_j(t))^2, i_0 = 1, 2, \dots, N, \quad (5)$$

where i_0 is the number of data sets, $\phi(t_{i_0})$ is the stochastic time effective function which can be referred to [17] given in Eq. (6), and T_j is the actual value from the output data sets.

$$\phi(t_{i_0}) = \frac{1}{\beta} \exp\left\{\int_{t_0}^{t_{i_0}} \mu(t)dt + \int_{t_0}^{t_{i_0}} \sigma(t)dB(t)\right\}, \quad (6)$$

where β is the time strength coefficient, t_0 is the time of the newest data in the data training set, and t_{i_0} is an arbitrary time point in the data training set. $\mu(t)$ is the drift function, $\sigma(t)$ is the volatility function, and $B(t)$ is the standard Brownian motion.

Then the total error of the output layer for average set in FM-ST-NN is defined as

$$\begin{aligned}
 l &= \frac{1}{N} \sum_{i_0=1}^N e_{i_0}(t) \\
 &= \frac{1}{N} \sum_{i_0=1}^N \phi(t_{i_0}) \frac{1}{2} \sum_{j=1}^p (T_j(t) - O_j(t))^2.
 \end{aligned}
 \tag{7}$$

To optimize the FM-ST-NN model, we often use the stochastic gradient decent method to update the weights until it achieves convergence. The algorithm of FM-ST-NN will be given in the following subsection.

2.2 The Algorithm of FM-ST-NN

The referred stochastic time effective function [17] which reflects the time random effects. When the events happened near, the investor will be affected heavily. The recent information has stronger effect than the old information does. It is used to modify the weights when training and finally predict the target more accurately. The training process of FM-ST-NN is detailed as follows:

(1). Choose five kinds of crude oil price related indexes as input data sets: daily opening price, daily highest price, daily lowest price, daily closing price and daily trade volume. Choose the closing price of crude oil price in the next trading day as the output (target) data sets. So there are five nodes in the input layer and one node in the output layer.

(2). Divide the sample data into training and testing data sets. Let the connective weights in vector \mathbf{v} and \mathbf{w} follow the uniform distribution on $(-1, 1)$.

(3). Give the stochastic time effective function value $\phi(t)$ according to some parameters. The transfer functions from the input layer to the hidden layer and the hidden layer to the output layer are both assumed as the sigmoid activation function.

(4). Predefined a minimum training threshold ε until the predicting result meet the standard rule. That is, if l is below the value ε , use the input data in the testing data sets to compute the prediction result (step 6). Otherwise, update the connective weights in the next step.

(5). Update the weights through the backward process and the goal is to minimize the value l . For the gradients of the weights in the output layer are computed as

$$\Delta w_{ij} = -\eta \frac{\partial l}{\partial w_{ij}}, i = 1, 2, \dots, m, j = 1, 2, \dots, p,
 \tag{8}$$

where η is the learning rate. So the update rule is

$$w_{ij} \leftarrow w_{ij} - \eta \frac{\partial l}{\partial w_{ij}}.$$

The gradients of the weights in the hidden layer are separately given by

$$\Delta v_{ij} = -\eta \frac{\partial l}{\partial v_{ij}}, i = 1, 2, \dots, n, j = 1, 2, \dots, m - k, \quad (9)$$

$$\Delta \hat{v}_{i,m-k+j} = -\eta \frac{\partial l}{\partial \hat{v}_{i,m-k+j}}, i = 1, 2, \dots, n, j = 1, 2, \dots, k. \quad (10)$$

So the update rule is

$$v_{ij} \leftarrow v_{ij} - \eta \frac{\partial l}{\partial v_{ij}};$$

$$\hat{v}_{i,m-k+j} \leftarrow \hat{v}_{i,m-k+j} - \eta \frac{\partial l}{\partial \hat{v}_{i,m-k+j}}.$$

(6). When the error l is below the predefined value ε , the predicting result for the output node p can be derived through

$$f\left[\sum_{j=1}^{m-k} w_{jp} f\left(\sum_{i=1}^n v_{ij} x_i\right) + \sum_{j=1}^k w_{m-k+j,p} f\left(\frac{1}{2} \left(\sum_{i=1}^n \hat{v}_{i,m-k+j} x_i\right)^2 - \sum_{i=1}^n (\hat{v}_{i,m-k+j})^2 x_i^2\right)\right].$$

3 Simulation Results and Evaluation

In this paper, we select the data of WTI crude oil from January 3, 2012 to December 31, 2018. At the beginning of modelling, all the data should be normalized. We use the formula

$$X' = \frac{X - \min X}{\max X - \min X} \quad (11)$$

to normalize the data in the range of $[0, 1]$. And it is easy to obtain the real value after predicting through converting as $X = X'(\max X - \min X) + \min X$.

We first use FM-ST-NN model predicting the crude oil price. The normalized data are divided into two parts: training data and testing data, with a proportion of 2:1. The number of neurons m in the hidden layer is 20 in the experiment and the maximum iteration number is 1000. The learning rate η is set as 0.03 and the minimum preset training threshold ε is 5×10^{-5} . As we considering the stochastic time effects on predicting, $\mu(t)$ and $\sigma(t)$ are assumed as

$$\begin{aligned} \beta &= 1, \\ \mu(t) &= \frac{1}{(b - t)^3}, \\ \sigma(t) &= 1. \end{aligned}$$

Figures 2–3 show the prediction results for training data sets and for testing data sets respectively.

Then we compare the predicting results from FM-ST-NN with different neural network models, i.e., NN and FM-NN. To analyze and evaluate the predicting performance of different models, the accuracy measures with the corresponding definitions are used below.

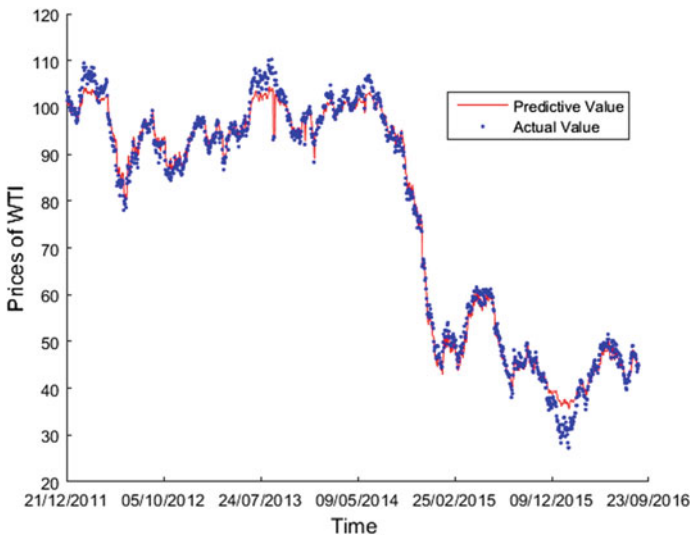


Fig. 2 The prediction results for training data sets ($k = 19$)

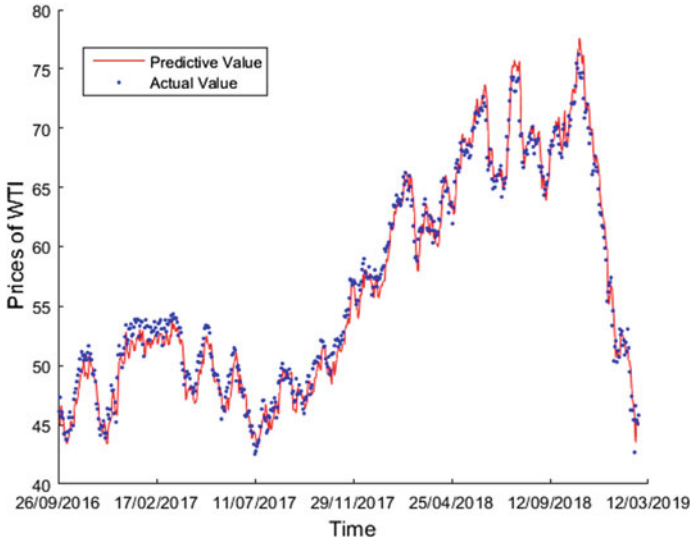


Fig. 3 The prediction results for testing data sets ($k = 19$)

$$MAE = \frac{1}{N} \sum_{i_0=1}^N |d_{i_0} - O_{i_0}|,$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i_0=1}^N (d_{i_0} - O_{i_0})^2},$$

$$MAPE = 100 \times \frac{1}{N} \sum_{i_0=1}^N \left| \frac{d_{i_0} - O_{i_0}}{d_{i_0}} \right|.$$

In order to know more clearly about the effects of factorization machine and stochastic time effect function, we evaluate the models through three accuracy measures listed in Table 1 and Table 2. Tables 1 and 2 give the performances of different predicting models, i.e., NN, FM-NN and FM-ST-NN on training data and testing data separately. From the two tables, we could find that both factorization machine technique and stochastic time effective function can improve the accuracies. And the FM-ST-NN model is the best among the other models on testing data. What’s more, deep interactions among the input nodes will improve the accuracies.

Table 1 The performances of different predicting models on training data

Models	Training data		
	MAE	RMSE	MAPE
NN	1.952	2.550	2.803
FM-NN ($k = 3$)	1.867	2.436	2.859
FM-ST-NN ($k = 3$)	1.946	2.525	2.883
FM-NN ($k = 19$)	1.548	2.069	2.322
FM-ST-NN ($k = 19$)	1.710	2.289	2.686

Table 2 The performances of different predicting models on testing data

Models	Testing data		
	MAE	RMSE	MAPE
NN	1.341	1.691	2.314
FM-NN ($k = 3$)	1.312	1.583	2.357
FM-ST-NN ($k = 3$)	1.269	1.551	2.259
FM-NN ($k = 19$)	1.094	1.367	1.926
FM-ST-NN ($k = 19$)	1.056	1.340	1.896

4 Conclusion

In this paper, the factorization machines based neural network combined with stochastic time effective function has been studied to predict the WTI crude oil price. We constructed the proposed neural network model and gave the algorithm to illustrate the method. After the experiments, we compared the forecasting results from the FM-ST-NN model with the other two neural network models including neural network and factorization machine based neural network. The aim is to find the advantages of the proposed model. According to the three accuracy measures, i.e., the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), FM-ST-NN model performs better. The innovation of this paper is combining factorization machine and stochastic time effect with neural network for predicting. On one way, the effect of stochastic time on the predicting can be explored deeply. And on the other way, we believe that combine factorization machine technology in various algorithm design will be a future work tackling problems arising from different applications.

References

1. Cheong, C. W. (2009). Modeling and forecasting crude oil markets using ARCH-typemodels. *Energy Policy*, 37, 2346–2355.
2. Franses, P. H., & Ghijssels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting*, 15, 1–9.

3. Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307–319.
4. Xie, W., Yu, L., Xu., S., & Wang., S. (2006) A new method for crude oil price forecasting based on support vector machines. In *International conference on computational science* (pp: 444-451). Springer-Verlag
5. Qian, X.Y. (2017). Financial series prediction: Comparison between precision of time series models and machine learning methods. *arXiv* : 1706.00948.
6. Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the taiwan stock index. *Computers and Operations Research*, 30, 901–923.
7. Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications*, 42, 3234–3241.
8. Tripathy, M. (2010). Power transformer differential protection using neural network principal component analysis and radial basis function neural network. *Simulation Modelling Practice and Theory*, 18, 600–611.
9. Ayadi, O. F., Williams, J., & Hyman, L. M. (2009). Fractional dynamic behavior in Forcados Oil Price Series: an application of detrended fluctuation analysis. *Energy for Sustainable Development*, 13, 11–17.
10. Cacciola, M., Megali, G., Pellicano, D., & Morabito, F. C. (2012). Elman neural networks for characterizing voids in welded strips: a study. *Neural Computing and Applications*, 21, 869–875.
11. Michael, Ford S. (2003). The illustrated wavelet transform handbook: introductory theory and applications in science. *J Phys A General Phys*, 37(2), 1947–1948.
12. Yousefi, S., Weinreich, I., & Reinartz, D. (2005). Wavelet-based prediction of oil prices. *Chaos, Solitons & Fractals*, 25, 265–275.
13. Armano, G., Marchesi, M., & Murru, A. (2005). A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, 170, 3–33.
14. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42, 2162–2172.
15. Wang, J., Pan, H. P., & Liu, F. J. (2012). Forecasting crude oil price and stock price by jump stochastic time effective neural network model. *Journal of Applied Mathematics*, 2012, 1–7.
16. Niu, H. L., & Wang, J. (2014). Financial time series prediction by a random data-time effective RBF neural network. *Soft Computing*, 18, 497–508.
17. Wang, J., & Wang, J. (2016). Forecasting energy market indices with recurrent neural networks: case study of crude oil price fluctuations. *Energy*, 102, 365–374.
18. Huang, L., & Wang, J. (2018). Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. *Energy*, 151, 875–888.
19. Rendle, S. (2010). Factorization machines. In *ICDM 2010, the IEEE international conference on data mining*, (pp. 995-1000), Sydney, Australia, 14-17 December
20. He, X., Zhang, H., Kan, M. Y., & Chua, T. S. (2017). Fast matrix factorization for online recommendation with implicit feedback. In *International ACM SIGIR Conference on Research and Development in Information Retrieval ACM*, (pp. 549–558), vol. (2016).
21. Liu, Y., Guo, W., Zang, D., & Li, Z. (2018) A hybrid neural network model with non-linear factorization machines for collaborative recommendation. *Information Retrieval*, 213-224
22. Wang, Q., Liu, F., Xing, S., Zhao, X., & Li, T. (2018). Research on CTR Prediction based on deep learning. *IEEE Access*, 7, 12779–12789.
23. Ni, W., Liu, T., Zeng, Q., Zhang, X., Duan, H., & Xie, N. (2018). Robust factorization machines for credit default Prediction. In *PRICAI 2018: Trends in Artificial Intelligence*, pp. 941–953.