A New Hybrid Approach for Analysis of Factors Affecting Crude Oil Price

Wei Xu[1], Jue Wang[2], Xun Zhang[2], Wen Zhang[1], and Shouyang Wang[1,2]

[1] School of Management, Graduate University of Chinese Academy of Sciences,
Chinese Academy of Sciences, Beijing, 100080, China
{xuw-06b1,zhangwen05}@mails.gucas.ac.cn
[2] Institute of Systems Science, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing, 100080, China
{wjue,zhangxun,sywang}@amss.ac.cn

Abstract. In this paper, a new hybrid approach is presented to analyze factors affecting crude oil price using rough set and wavelet neural network. Related factors that affect crude oil price are found using text mining technique and Brent oil price is chosen as the decision price because it plays an important role in world crude oil markets. The relevant subsets of the factors are discovered by rough set module and the main factors are got, and then the important degrees of these are measured using wavelet neural network. Based on the novel hybrid approach, the predictability of crude oil price is discussed.

Keywords: crude oil price, rough set, wavelet neural network, prediction.

1   Introduction

The high volatility and irregularity of crude oil market creates uncertainty, mainly because of the interaction of many factors in crude oil markets. How to analyze and use these factors to forecast the crude oil price has attracted increasing attention from academics and practitioners in the past decade. The literature on factors affecting crude oil price relates oil shocks either to the instability of the market structures or to the effect of the price elasticity of demand [1-4]. Mork [1] and Huntington [2] demonstrated the asymmetric relationship that a reduction in oil prices does not necessarily lead to noticeable output growth, while an increase can have a negative impact on output growth. The study of Ferderer points to the observation that disruptions in oil market not only give rise to higher prices, but also increase oil price volatility [3]. C.W.Yang [4] analyze factors affecting price volatility of the US oil market by examining the market structure of OPEC, the stable and unstable demand structure, and related elasticity of demand.

However, there are still several unsolved issues with the above methods. The first problem is that only a few factors, such as demand and supply, are taken into account in the researchers' models. In fact, many other factors have a combined effect on crude oil prices: economic, military, natural disasters and speculation, as well as people's expectations. These important factors can be hard to handle and they are not

included in the traditional models. The second problem is how to quantify the qualitative factors. Some qualitative variables, such as political variables, are not easy to quantify due to uncertainty. Furthermore, related data collection is very difficult. Thus, the above proposed models are not easy to operate or impractical. Therefore, it is important that new methods be developed for factors analysis.

Considering the discussed above, we present a new hybrid approach for factors affecting oil price analysis based on rough set (RS) and wavelet neural network (WNN). The related factors that affect oil price are found using text mining technique. The relevant subsets of the factors are discovered by rough set module and the main factors are got. The important degrees of these are measured using wavelet neural network. Based on the novel hybrid approach, the predictability of crude oil price is discussed. Finally, we draw the conclusions.

2   A Hybrid Approach Based on RS and WNN

2.1   Basic Concepts of Rough Set Theory

The theory of rough set (RS), proposed by Pawlak [5-6], has been proved to be a powerful tool for handling various types of data. Several models based on original rough set theory have been presented to solve different problems [7-8]. It has important applications to artificial intelligence and economic and financial prediction [9-10], as a tool for dealing with vagueness and uncertainty of facts, and in classification.

The set about objects is represented in the form of an information system. The rows of the information system are labeled by objects, whereas columns are labeled by attributes and entries of the information system are attribute values.

Formally, an information system (IS) is an quadruple $\varphi = (U, AT, V, f)$, where U is a finite nonempty set of objects and AT is a finite nonempty set of attributes, $V= \cup_{a \in AT} V_a$ and V_a is a domain of attribute a, f: $U \times AT \rightarrow V$ is a total function such that $f(x, a) \in V_a$ for every $x \in U$, $a \in AT$, called an information function. An IS can be seen as decision system assuming attributes concluding two part $(AT \quad D)$, and $AT \cap D = \phi$, where AT is called condition attributes and D contains decision attributes.

Let $R_A= \{(x, y) \in U \times U: f_a(x) = f_a(y), \forall a \in A\}$. The relation R_A thus partitions the universe of discourse U into disjoint subsets called equivalence classes, and the result is denoted by U/A which forms a partition of U. The equivalence class including x is denoted by $[x]_A$, that is, $[x]_A = \{y \in U: (x, y) \in R_A\}$. The sets in the partition U/A are referred to as A-elementary sets. Objects from $[x]_A$ are indiscernible with regard to their descriptors in the system. Given an arbitrary set $X \subseteq U$, it may be impossible to describe X precisely using the equivalence classes in U/A. In this case, one may characterize X by a pair of lower and upper approximations:

$$\underline{A}(X) = \{x \in U : [x]_A \subseteq X\} \tag{1}$$

$$\overline{A}(X) = \{x \in U : [x]_A \cap X \neq \phi\} \tag{2}$$

The lower approximation $\underline{A}(X)$ is the set of objects that belong to X with certainty, while the upper approximation $\overline{A}(X)$ is the set of objects that possibly belong to X.

In an IS, if $R_B = R_{AT}$, $B \subseteq AT$, then B is referred to as a consistent set of φ. If $B \subseteq AT$ is a consistent set of φ, and no proper subset of B is a consistent set of φ, then B is referred to as a reduction of φ. Similarly, In decision system, $\varphi = (U, AT, V, f, D)$, satisfying $R_{AT} \subseteq R_D$, if $R_B \subseteq R_D$, $B \subseteq AT$, then B is referred to as a consistent set of φ. If $B \subseteq AT$ is a consistent set of φ, and no proper subset of B is a consistent set of φ, then B is referred to as a reduction of φ.

2.2 Basic Concepts of Wavelet Neural Network

Wavelet neural network (WNN) is a novel approach towards the learning function [11]. Wavelet networks, which combine the wavelet theory and feed-forward neural networks, utilize wavelets as the basis function to construct a network. Wavelet function is a local function and influences the networks' output only in some local ranges. The wavelet neural network shows surprising effectiveness in solving the conventional problems of poor convergence or even divergence encountered in other kinds of neural networks [12]. The WNN consists of three layers: input layer, hidden layer and output layer. The connections between input-hidden units and hidden-output units are called weights u_{ti} and w_t, respectively. A Morlet mother function is used as node activation function for the hidden layer. The dilation and translation parameters, a_t and b_t, of the Morlet function for each node in the hidden layer are different and they need to be optimized. In the WNN, the gradient descend algorithm is employed and the error is minimized by adjusting u_{ti}, w_t, a_t and b_t parameters [13]. In the WNN, the following steps are carried out:

Step1. Initializing the dilation parameter a_t, translation parameter b_t and node connection weights u_{ti}, w_t to some random values. All those random values are limited in the interval $(0, 1)$.

Step2. Inputting data $x_n(i)$ and corresponding output values v_n^T, where the superscript T represents the target output state.

Step3. Propagating the initial signal forward through the network:

$$v_n^T = \sum_{t=1}^{T} w_t h\left(\frac{\sum_{i=1}^{s} u_{ti} x_n(i) - b_t}{a_t} \right) \tag{3}$$

Where h is taken as a Morlet wavelet

$$h(t) = \cos(1.75t) \exp\left(-\frac{t^2}{2} \right) \tag{4}$$

Step4. Calculation of the WNN parameters: $\Delta w_t = -\eta \dfrac{\partial E}{\partial w_t} + \alpha \Delta w_t$, $\Delta u_{ti} = -\eta \dfrac{\partial E}{\partial u_{ti}} + \alpha \Delta u_{ti}$,

$\Delta a_t = -\eta \dfrac{\partial E}{\partial a_t} + \alpha \Delta a_t$, $\Delta b_t = -\eta \dfrac{\partial E}{\partial b_t} + \alpha \Delta b_t$, the error function E is taken as $E = \dfrac{1}{2} \sum_{n=1}^{N} \left(v_n^T - v_n \right)^2$,
and v_n^T, v_n are the experimental and calculated values, respectively. N stands for the data number of training set, and η and α being the learning rate and the momentum term, respectively.

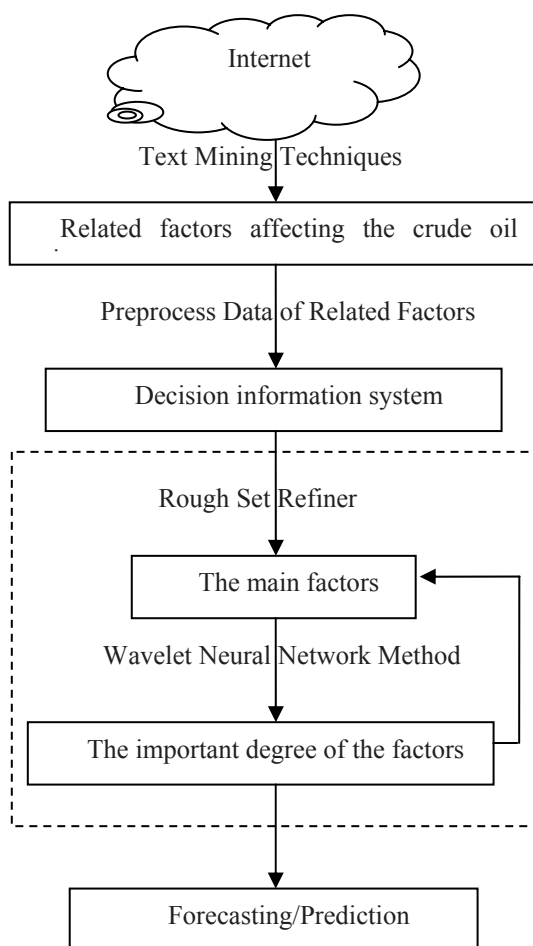Step5. The WNN parameters were changed until the network output satisfies the error criteria.

Fig. 1. The Main Process of the Proposed Hybrid Approach

2.3   A Hybrid Approach Based on RS and WNN

As revealed in previous two subsections, rough set theory is a powerful mathematical tool that handles vagueness and uncertainty. It can reduce the attributes from the large database and get the most important, but it can't find the important degree from the reduced attributes. However, wavelet neural network can did a good work on getting the important degree of the input, but its input variables are limited because of the algorithm's speed and astringency. In view of this point, a new hybrid approach— RSWNN approach— is proposed for factors affecting oil price analysis.

The proposed approach consists of three modules: the text mining module, the rough set module and the wavelet neural module. In the first module, as the preprocessing text mining techniques are used to mine the related factors affecting the oil price. The text mining theory is a powerful technique which draws on information retrieval, data mining, machine learning, statistics and computational linguistics [14-15]. The text mining process consists of four major stages: collecting documents, preprocessing documents, feature extraction, and metadata mining. The details have been given in Ref. [16]. In the second module, rough set theory is used as a refiner for the factors mining from the text mining technique. In the third module, WNN are used to measure the important degree of the factors getting from the rough set refiner. Based on the proposed approach, the predictability of crude oil price is discussed. Generally, the process of the proposed hybrid approach is illustrated in Fig. 1.

As can be seen from Fig. 1, the novel hybrid approach is actually an intelligent analysis system. Text mining techniques are used to mine the related factors affecting the crude oil price automatically. Decision information system is got through preprocessing the data of the related factors. And then as the main body the rough set refiner is used to find the main factors from a large number of factors affecting the crude oil price while the wavelet neural network are used to measure the important degree of the factors getting from the rough set refiner. Finally, based on the proposed approach, the predictability of crude oil price is discussed.

3   An Application of the Hybrid Approach

3.1   Presentation of the Text Mining Results and the Data

In Section 2.3, the text mining stages are described. In this section, the factors called attributes or metadata are presented in conjunction with crude oil price. Then "crude oil price, factor", "crude oil market, factor" and "crude oil volatility, factor" are used as keywords to search related data, including some numeric and textual data. We obtain the following factors affecting crude oil price, as shown in Table 1.

We select annual data of these attributes because of analyzing the long-run factors affecting oil price. The sources of these data are in the EIA, Reuters and IFS and the period of the data begin from 1970, and end of 2005. We choose Brent oil price as the decision attribute d because it plays an important role in world crude oil markets.

Table 1. The Factors Affecting Crude Oil Price

No	Factors	No	Factors	No	Factors
a_1	world total demand	a_{13}	Core OPEC production capacity	a_{25}	OPEC oil embargo
a_2	world total supply	a_{14}	Non-OPEC production	a_{26}	economic sanction to oil nations
a_3	China/U.S. foreign exchange rate	a_{15}	Non-OPEC production capacity	a_{27}	large oil company merger
a_4	Japan / U.S. foreign exchange rate	a_{16}	fuel switching capacity	a_{28}	rumors and false news
a_5	France / U.S. foreign exchange rate	a_{17}	OECD total stocks	a_{29}	oil worker strike
a_6	U.S. / U.K foreign exchange rate	a_{18}	world total reserves	a_{30}	hostage crisis
a_7	world economy growth	a_{19}	the distance in wells	a_{31}	terrorist attack
a_8	speculation	a_{20}	the number of drilling wells	a_{32}	political conflict in oil nations
a_9	OPEC market share	a_{21}	the number of oil wells	a_{33}	revolutions in oil countries
a_{10}	forward price of crude oil	a_{22}	gasoline tax rate	a_{34}	wars in oil countries
a_{11}	OPEC production	a_{23}	oil import fee	a_{35}	geopolitics
a_{12}	Core OPEC production	a_{24}	CO_2 limit	a_{36}	natural disasters related to oil

3.2 Presentation of the Rough Set Results

The above forms an information table $\varphi = (U, AT, V, f, D)$. The continuous attributes are discredited by setting norms dividing the original domains of the attributes into sub-intervals using equal frequency method.

The rough set analysis of the IS has been performed using an attribute reductions algorithm with genetic algorithm. 85 reductions were obtained based on the IS. They contain 5-8 attributes, which are considerably smaller than 36-the total number of attributes. This result gives the idea of reduction a strong support because each of the reductions contains fewer attributes, but, on the other hand, ensures the same value of quality of approximation as the whole set of attributes AT.

The proper reduction is selected among these reductions. The selection is based on two criteria: the reduction should contain as small a number of attributes as possible, and the reduction should not miss the attributes judged by the experts as the most significant factors. As a result, the reduction selected, as the main factors affecting the crude oil price, was the #31, which includes: a_1 (world total demand), a_2 (world total supply), a_4 (Japan / U.S. foreign exchange rate), a_6 (U.S. / U.K foreign exchange rate), a_7 (world economy growth) and a_{17} (OECD total stocks).

3.3 Presentation of the Wavelet Neural Network Results

In this subsection the important degree of the main factors - a_1 (world total demand), a_2 (world total supply), a_4 (Japan / U.S. foreign exchange rate), a_6 (U.S. / U.K foreign exchange rate), a_7 (world economy growth) and a_{17} (OECD total stocks) is got using wavelet neural network. In this module, the input variables are the data of these main factors and the output variable is the Brent oil price.

   The wavelet neural network is constructed based on these selected variables. In order to improve the algorithm stability and quicken the training, the input variables and the output variables are normalized respectively before training the WNN so that the data of learning samples are varied within small bound. The epoch is stopped when MSE satisfied the given error criterion. As a result, the selected input variables with their important degree in parentheses were world total demand (0.2021), world total supply (0.1745), Japan / U.S. foreign exchange rate (0.1953), U.S. / U.K foreign exchange rate (0.1011), world economy growth (0.1672) and OECD Total Stocks (0.1607). The important degree of the input variables describes the influence of these factors to the crude oil price.

   As can be seen above, world total demand and supply are the most important factors affecting crude oil price, and relatively demand is dominant. Crude oil stock can reduce the volatility of crude oil price. World economy growth and exchange rates also affect the crude oil price and increase the volatility of crude oil price.

3.4 The Predictability of the Proposed Approach

In this subsection, we discuss the predictability of the proposed approach. The annual data of these main factors and Brent oil price in 1970-2002 are used as training data, and the annual data in 2003-2005 are used as testing data. The result shows that MSE of the testing data is 2.94%. The hybrid approach can be used as a tool for crude oil price forecasting and improve the long-run prediction quality in view of empirical results.

4 Conclusions

In this paper, a new hybrid approach is presented to analyze factors affecting crude oil price using rough set and wavelet neural network. World total demand, world total supply, Japan / U.S. foreign exchange rate, U.S. / Euro foreign exchange rate, world economy growth and OECD total stocks are the main factors, and the important degree are found. In future work we will construct an intelligent system based on the proposed approach for forecasting crude oil price and compared with other models.

References

1. Mork K.A., Oil and the Macroeconomy When Prices Go Up and Down: an Extension of Hamilton's Results. Journal of Political Economy, 97 (1989) 740-744
2. Huntington H.G., Crude Oil Prices and US Economic Performance: Where Does The Asymmetry Reside? The Energy Journal, 19(4) (1998) 107-132

3. Ferderer J.P., Oil Price Volatility and the Macroeconomy: a Solution to the Asymmetry Puzzle. Journal of Macroeconomics, 18 (1996) 1-16

4. C.W. Yang, M.J. Hwang and B.N. Huang, An Analysis of Factors Affecting Price Volatility of the US Oil Market. Energy Economics, 24 (2002) 107-119

5. Pawlak Z., Rough Set. International Journal of Computer and Information Sciences, 11(5) (1982) 341-356

6. Pawlak Z., Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)

7. Yao Y.Y., Generalized Rough Set Model, In Rough Sets in Knowledge Discovery1: Methodology and Applications, L. Polkowski and A. Skowron (eds), Heidelberg: Physica-Verlag (1998)

8. Wezhi Wu, Wenxiu Zhang, Huaizu Li, Knowledge Acquisition in Incomplete Fuzzy Information Systems via the Rough Set Approach. Expert Systems, 20(5) (2003) 280-286

9. Slowinski R., Zopounidis C., Dimitras A., Prediction of Company Acquisition in Greece by Means of the Rough Set Approach. European Journal of Operational Research, 100 (1997) 1-15

10. Tay F.E. and Lixiang Shen, Economic and Financial Prediction Using Rough Sets Model. European Journal of Operational Research, 141 (2002) 641-659

11. Q. Zhang and A. Benveniste, Wavelet Networks. IEEE Trans. Neural Network, 3(6) (1992) 889–898

12. Zhang X, Qi J, Zhang R, Liu M, Hu Z, Xue H, et al. Prediction of Programmed-temperature Retention Values of Naphtha by Wavelet Neural Network. Computers and Chemistry, 25 (2) (2001) 125-133

13. Tabaraki R, Khayamian T, Ensafi A. A., Principle Component-wavelet Neural Networks as a New Multivariate Calibration Method. Analytical Letters, 38 (2005) 1-13

14. R. Chau, C.H. Yeh, A Multilingual Text Mining Approach to Web Cross-lingual Text Retrieval. Knowledge-Based Systems, 17, (5-6) (2004) 219-227

15. A. Gelbukh, Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science, Vol. 2004. Springer-Verlag, Berlin Heidelberg (2001)

16. Lean Yu, Shouyang Wang, K.K. Lai, A Rough-Set-Refined Text Mining Approach for Crude Oil Market Tendency Forecasting. International Journal of Knowledge and Systems Sciences, 2(1) (2005) 33-46