Oil Price Forecasting with an EMD-Based Multiscale Neural Network Learning Paradigm

Lean Yu^{1,2}, Kin Keung Lai², Shouyang Wang¹, and Kaijian He²

¹ Institute of Systems Science, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100080, China
{yulean, sywang}@amss.ac.cn

² Department of Management Sciences, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong
{msyulean, mskklai, paulhekj}@cityu.edu.hk

Abstract. In this study, a multiscale neural network learning paradigm based on empirical mode decomposition (EMD) is proposed for crude oil price prediction. In this learning paradigm, the original price series are first decomposed into various independent intrinsic mode components (IMCs) with a range of frequency scales. Then the internal correlation structures of different IMCs are explored by neural network model. With the neural network weights, some important IMCs are selected as final neural network inputs and some unimportant IMCs that are of little use in the mapping of input to output are discarded. Finally, the selected IMCs are input into another neural network model for prediction purpose. For verification, the proposed multiscale neural network learning paradigm is applied to a typical crude oil price — West Texas Intermediate (WTI) crude oil spot price prediction.

Keywords: Crude oil price forecasting, artificial neural networks, empirical mode decomposition, multiscale learning paradigm.

1 Introduction

As is known to all, crude oil price forecasting is a rather challenging task due to high volatility and effects of irregular events. In the past practices, traditional statistical and econometric techniques are widely applied to crude oil price forecasting. For example, Huntington [1] applied a sophisticated econometric model to predict crude oil prices in the 1980s. Abramson and Finizza [2] utilized a probabilistic model for predicting oil prices, and Morana [3] suggested a semi-parametric statistical method for short-term oil prices forecasting. Usually, these models can provide good results when the time series under study is linear or near linear. However, in real-world crude oil price series, there is a great deal of highly nonlinearity and irregularity. Numerous experiments [4-5] have demonstrated that the performance might be very poor if one continued using these traditional statistical and econometric models. The main reason leading to this phenomenon is that the traditional statistical and econometric models are built on the linear assumptions and they cannot capture the nonlinear patterns hidden in the time series.

With the advent of artificial intelligence (AI), neural networks provide a powerful alternative solution to nonlinear time series prediction. In this situation, using neural network as a nonlinear modeling technique for nonlinear crude oil price forecasting has been a common practice in the past decades [6-7]. Unfortunately, the currently employed neural network learning process, both the training algorithms and preprocessing methods used, is essentially a single-scale learning procedure. That is, it is trained based on a pure time series. A possible drawback of such a conventional neural network learning process is that it is sometimes inadequate for complex and difficult problems [8], e.g., crude oil price forecasting problems, and often yields poor generalization performance. To alleviate the potential problem of the poor generalization, a multiscale neural network learning paradigm based on empirical mode decomposition (EMD) is proposed. In the multiscale learning paradigm, the original time series are first decomposed into various independent intrinsic mode components (IMCs) with a range of frequency scales. Then the internal correlation structures of different IMCs are explored by neural network. With the neural network weights, some important IMCs are selected as final neural network inputs and some unimportant IMCs that are of little use in the mapping of input to output are discarded. Finally, the selected IMCs are input into another neural network model for prediction purpose.

The rest of this study is organized as follows. Section 2 describes the building process of the proposed EMD-based multiscale neural network learning paradigm in detail. For verification, a typical crude oil price — West Texas Intermediate (WTI) crude oil spot price is used in Section 3. And Section 4 concludes the article.

2 Multiscale Neural Network Learning Paradigm

In this section, an overall formulation process of the EMD-based multiscale neural network learning paradigm is proposed for nonlinear time series prediction such as crude oil price forecasting. First of all, the EMD technique is briefly reviewed. Then the EMD-based multiscale neural network learning paradigm is proposed.

2.1 Empirical Mode Decomposition (EMD)

The empirical mode decomposition (EMD) method first proposed by Huang et al. [9] is a form of adaptive time series decomposition technique using spectral analysis via Hilbert transform for nonlinear and nonstationary time series data. Traditional forms of spectral analysis, like Fourier, assume that a time series (either linear or nonlinear) can be decomposed into a set of linear components. As the degree of nonlinearity and nonstationarity in a time series increases, the Fourier decomposition often produces large sets of physically meaningless harmonics when applied to nonlinear time series [10]. For wavelet analysis, it needs to select a filter function beforehand [11], which is difficult for some unknown time series. Naturally, a new spectrum analysis method, EMD based on Hilbert transform, is emerged.

The basic principle of EMD is to decompose a time series into a sum of intrinsic mode components (IMCs) with the following sifting procedure.

- (1) Identify all the local extrema including local maxima and minima of $x(t)$,
- (2) Connect all local extrema by a cubic spline line to generate its upper and lower envelopes $x_{up}(t)$ and $x_{low}(t)$.
- (3) Compute the point-by-point envelope mean $m(t)$ from upper and lower envelopes, i.e., $m(t) = (x_{up}(t) + x_{low}(t))/2$.
- (4) Extract the details, $d(t) = x(t) - m(t)$.
- (5) Check the properties of $d(t)$: (i) if $d(t)$ meets the above two requirements, an IMC is derived and replace $x(t)$ with the residual $r(t) = x(t) - d(t)$; (ii) if $d(t)$ is not an IMC, replace $x(t)$ with $d(t)$.
- (6) Repeat Step 1) – 5) until the residual satisfies the following stopping condition: $\sum_{t=1}^T \frac{[d_j(t) - d_{j+1}(t)]^2}{d_j^2(t)} < SC$, where $d_j(t)$ is the sifting result in the j th iteration, and SC is the stopping condition. Typically, it is usually set between 0.2 and 0.3.

The EMD extracts the next IMC by applying the above procedure to the residual term $r_1(t) = x(t) - c_1(t)$, where $c_1(t)$ denotes the first IMC. The decomposition process can be repeated until the last residue $r(t)$ only has at most one local extremum or becomes a monotonic function from which no more IMCs can be extracted. A typical sifting process can be represented by a tree graph, as illustrated in Fig. 1.

Fig. 1. The tree graph representation of the EMD sifting process

At the end of this sifting procedure, the time series $x(t)$ can be expressed by

$$x(t) = \sum_{j=1}^p c_j(t) + r_p(t) \quad (1)$$

where p is the number of IMCs, $r_p(t)$ is the final residue, which is the main trend of $x(t)$, and $c_j(t)$ ($j = 1, 2, \dots, p$) are the IMCs, which are nearly orthogonal to each other, and all have nearly zero means. Thus, one can achieve a decomposition of the data series into m -empirical modes and a residue. The frequency components contained in each frequency band are different and they change with the variation of time series $x(t)$, while $r_p(t)$ represents the central tendency of time series $x(t)$. For later analysis, the residue has been seen as the $(p+1)$ th IMC (i.e., c_{p+1}). Then the Eq. (1) can be rewritten as

$$x(t) = \sum_{j=1}^{p+1} c_j(t) \quad (2)$$

2.2 EMD-Based Multiscale Neural Network Learning Paradigm

As we know, artificial neural networks (ANNs) are a new kind of intelligent learning algorithm and are widely used in some application domains. In this study, a standard three-layer feed-forward neural network (FNN) [12] is selected as a multiscale neural network learning tool for nonlinear time series prediction. The main reason of selecting FNN as a predictor is that an FNN is often viewed as a “universal approximator” [12]. However, a major challenge in neural network learning is how to make the trained networks possess good generalization ability, i.e., they can generalize well to cases that were not included in the training set. Some researchers argued to use the cross-validation technique for getting good generalization [13]. But in the cross validation technique, neural network learning is based on a single series representation for the entire training process. However, when the problem is very difficult and complex, single series representation for neural network learning may be inadequate [8]. For this reason, the EMD-based multiscale neural network learning is employed to decompose a time series and approximating it using decomposed components via a multi-variable analysis framework. Generally speaking, the EMD-based multiscale neural network learning paradigm consists of three different steps.

Step I: Decomposition. The original time series are decomposed into various independent intrinsic mode components (IMCs) with a range of frequency scales. These produced different IMCs can formulate the inputs of neural networks. Fig. 2 presents an illustrative decomposition example of a time series with p IMCs and one residue.

Fig. 2. The EMD decomposition to form the inputs of the neural networks

Step II: Selection. The IMCs produced by Step I are input into the neural networks for training. In the neural network learning process, the internal correlation structures of different IMCs should be explored. Using the explored internal relationships, some important IMCs are selected as final neural network inputs and some unimportant IMCs that are of little use in the mapping of input to output are discarded. The mapping is expected to be highly nonlinear and dependent on the characteristics of the individual time series data. In this study, we use the connection weights of neural networks as a selection criterion. Let

$$s_i = \sum_{j=1}^q |w_{ij}| \quad (3)$$

represent the importance of input c_i (e.g., IMC_i), where w_{ij} is the weight of input neuron i to hidden neuron j and q is the number of hidden nodes.

Subsequently, we use the following relative importance of input c_i (e.g., IMC_i) as a final selection criterion, i.e.,

$$\tilde{s}_i = s_i / \sum_{i=1}^p s_i \quad (4)$$

where $\tilde{s}_i$ represents the normalized input strength and p is the number of input nodes.

Usually, the input components (i.e., IMC_i) with small $\tilde{s}_i$ will be considered to be unimportant and may be discarded without affecting the prediction performance. Of course, as an alternative solution, principal component analysis (PCA) [14] can also be used to explore the relationships of different inputs, but it cannot effectively capture internal relationship between dependent variable and decomposed components.

Step III: Prediction. The final step is to use the selected IMCs with larger $\tilde{s}_i$ to train another neural network model for nonlinear time series prediction purpose.

It is interesting to examine the underlying idea of the multiscale learning paradigm. For a complex time series problem, a neural network is inadequate to approximate it due to lack of internal correlations. Through the EMD decomposition, the internal correlation structures at different scale levels that may be obscured in the original series are exposed by neural network learning. Usually, how well a network learns the internal correlation structure influences the degree of neural network generalization [8]. When the internal relationship is explicitly revealed to the neural networks, it can be more easily captured and learned.

3 Experiment

For evaluation and verification, a typical crude oil price series, West Texas Intermediate (WTI) crude oil spot price is chosen as the experiment sample data since the WTI price is most famous benchmark prices used widely as the basis of many crude oil price formulas. The WTI oil price data used here are daily data and are obtained from the website of Department of Energy (DOE) of US (<http://www.eia.doe.gov/>). We take the daily data from January 1, 1998 to October 30, 2006 excluding public holidays with a total of 2210 observations. Particularly, the data from January 1, 1998 till December 31, 2004 is used for training set (1751 observations) and the remainder is used as testing set (459 observations).

For performance comparison, the normalized mean squared error (NMSE) is used as the evaluation criterion, which can be represented by

$$NMSE = \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \bar{x}_i)^2} = \frac{1}{\sigma^2} \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (5)$$

where σ^2 is the estimated variance of the testing data, x_i and $\hat{x}_i$ are the actual and predicted value, $\bar{x}_i$ being the mean, and N is the number of testing data. If the estimated mean of the data is used as predictor, $NMSE=1.0$ is obtained. The $NMSE$ is

related to R^2 which measures the dependence between pairs of desired values and predictions by $NMSE = 1 - R^2$. For comparison purpose, the conventional single-scale neural network learning paradigm is used as the benchmark model.

Following the multiscale learning paradigm, the WTI training data are decomposed into eight IMCs and one residue, as shown in Fig. 3. From this decomposition, a neural network model is used to select some important components for final multiscale learning. We use neural network with the architecture (9:15:1) to perform this selection task. That is, the neural network has nine inputs, fifteen hidden nodes and one

Fig. 3. The EMD decomposition of WTI crude oil spot price

Fig. 4. The distribution of relative importance indicator of eleven IMCs

output neuron. Note that the number of hidden nodes is determined by trial and error and practical problems. Fig. 4 illustrates the normalized input strength or relative importance indicator of all decomposed components.

As can be seen from Fig. 4, it is not hard to find that the input components IMC1, IMC2 and IMC6 are consistently less important than other inputs. In this way, the remaining six IMC components are finally chosen as neural network inputs for final prediction purpose. With the decomposed IMCs, some simulations are re-implemented after these less important inputs are eliminated. This leads to a new neural network architecture (6:15:1). In this study, we represent the multiscale neural network by $(6:15:1) = (9:15:1) - [1, 2, 6]$. For consistency, we still use 15 hidden neurons. Table 1 shows the crude oil price prediction performance with different neural network learning paradigms.

Table 1. The out-of-sample prediction results for different neural network learning paradigms

Neural Network Model		Prediction Performance		
Architecture	Learning type	NMSE	Minimum	R ²
(9:15:1)	Single-scale	0.0712	0.0102	0.9288
	Multiscale	0.0143	0.0082	0.9767
(6:15:1)	Single-scale	0.0587	0.0095	0.9413
	Multiscale	0.0084	0.0008	0.9876

As can be revealed from Table 1, several interesting conclusions can be found. First of all, the multiscale learning generally performs better than the single scale learning in terms of different neural network architectures. Second, in the same network architecture, the performance of the multiscale learning is much better than that of the single scale learning. Third, when the network architecture is changed from (9:15:1) to (6:15:1), the single scale neural network learning model with architecture (6:15:1) can outperform another single scale neural network learning model with architecture (9:15:1). The main reason is that the crude oil prices are more dependent on some nearer crude oil price value. If we use some over-length lag term as neural network inputs, the performance may be worse because much redundancy is involved. Finally, after some less important inputs are eliminated, the multiscale neural network learning can show a more consistent performance for different weight initializations than the conventional single scale neural network learning.

4 Conclusions

In this study, an EMD-based multiscale neural network learning paradigm is proposed for complex nonlinear time series prediction problem. In this proposed multiscale learning paradigm, the original time series are first decomposed into various independent intrinsic mode components (IMCs) with different scales. Then these IMCs are input into neural network for exploring the internal correlation structures of different IMCs. Using the relative importance indicator, some important IMCs are selected as final neural network inputs and some unimportant IMCs that are of little use in the mapping of input to output are discarded. Finally, the retained IMCs are input into another neural network model for final prediction purpose. For verification

purpose, a typical crude oil price, WTI spot price series is used. Experimental results obtained confirm that the multiscale neural network learning paradigm can effectively improve the generalization capability, implying that the proposed multiscale neural network learning paradigm can be used as a promising tool for complex nonlinear time series prediction such as crude oil price forecasting problem.

Acknowledgements

This work is supported by the grants from the National Natural Science Foundation of China (NSFC No. 70221001, 70601029), the Chinese Academy of Sciences (CAS No. 3547600), the Academy of Mathematics and Systems Sciences (AMSS No. 3543500) of CAS, and the Strategic Research Grant of City University of Hong Kong (SRG No. 7001677, 7001806).

References

1. Huntington, H.G.: Oil Price Forecasting in the 1980s: What Went Wrong? *The Energy Journal* 15(2) (1994) 1–22
2. Abramson, B., Finizza, A.: Probabilistic Forecasts from Probabilistic Models: A Case Study in the Oil Market. *International Journal of Forecasting* 11(1) (1995) 63–72
3. Morana, C.: A Semiparametric Approach to Short-term Oil Price Forecasting. *Energy Economics* 23(3) (2001) 325–338
4. Cichocki, A., Unbehauen, R.: *Neural Networks for Optimization and Signal Processing*. Wiley, New York (1993)
5. Weigend, A.S., Gershenfeld, N.A.: *Time Series Prediction: Forecasting the Future and Understanding the Past*. Addison-Wesley, Reading, MA (1994)
6. Wang, S.Y., Yu, L., Lai, K.K.: A Novel Hybrid AI System Framework for Crude Oil Price Forecasting. *Lecture Notes in Artificial Intelligence* 3327 (2004) 233–242
7. Wang, S.Y., Yu, L., Lai, K.K.: Crude Oil Price Forecasting with TEI@I Methodology. *Journal of Systems Science and Complexity* 18(2) (2005) 145–166
8. Liang, Y., Page, E.W.: Multiresolution Learning Paradigm and Signal Prediction. *IEEE Transactions on Signal Processing* 45 (1997) 2858–2864
9. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences* 454 (1998) 903–995
10. Huang, N.E., Shen, Z., Long, S.R.: A New View of Nonlinear Water Waves: The Hilbert Spectrum. *Annual Review of Fluid Mechanics* 31 (1999) 417–457
11. Li, X.: Temporal Structure of Neuronal Population Oscillations with Empirical Mode Decomposition. *Physics Letters A* 356 (2006) 237–241
12. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2 (1989) 359–366
13. Krogh, A., Vedelsby, J.: Neural Network Ensembles, Cross Validation, and Active Learning. In Tesauro, G., Touretzky, D., Leen, T. (eds.): *Advances in Neural Information Processing Systems*, Cambridge, MA, MIT Press 7 (1995) 231–238
14. Yu, L., Wang, S.Y., Lai, K.K.: A Novel Nonlinear Ensemble Forecasting Model Incorporating GLAR and ANN for Foreign Exchange Rates. *Computers & Operations Research* 32 (2005) 2523–2541