INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 0975-4520

Prediction of Crude Oil Prices using Support Vector Regression

KOVVURI LAKSHMI PRIYA DARSHINI
PG Scholar, Department of Computer Science,
SVKP & Dr K S Raju Arts & Science College,
Penugonda, W.G.Dt., A.P, India
priyadarashini99@gmail.com

CHIRAPARAPU SRINIVASA RAO*
Associate Professor in Computer Science,
SVKP & Dr K S Raju Arts & Science College,
Penugonda, W.G.Dt., A.P, India
chiraparapu@gmail.com

Abstract:

The aim of this research is forecasting crude oil prices using Support Vector Regression (SVR). Algorithm to determine the optimal parameters in the model using the SVR is a grid search algorithm. This algorithm divides the range of parameters to be optimized into the grid and across all points to get the optimal parameters. In its application the grid search algorithm should be guided by a number of performance metrics, usually measured by cross-validation on the training data. Therefore, it is advisable to try some variations pair hyperplane parameters on SVR. Based on analysis calculation of accuracy and the prediction error using the training data generating R2 99.10868% while the value of MAPE by 1.789873%. The data testing generates R2 96.1639% while the value of MAPE by 1.942517%. This indicates to the data of testing using a linear kernel or accuracy of prediction accuracy results are quite large. Best model using the SVR has been formed can be used as a predictive model of crude oil prices. The results obtained showed crude oil prices from period 1 up to 10 experiencing decline.

Keywords: Crude Oil Prices, SVR, Kernel, Grid Search, Cross Validation

1. INTRODUCTION

Final energy consumption in Indonesia for the period 2000–2012 increased by an average of 2.9% per year. The most dominant type of energy is petroleum products which include aviation fuel, avgas, gasoline, kerosene, diesel oil, and fuel oil. These types of fuel consumed mostly by the transport sector. Today, most of the fuel prices are still subsidized. Fuel subsidies in 2013 have reached 199 trillion rupiahs. The government is also still subsidizing electricity for a particular type of users. Total electricity subsidies in 2013 reached 100 trillion rupiahs. The energy subsidy (fuel and electricity) has been increasing steadily. Energy subsidies in 2011 amounted to 195.3 trillion rupiahs and increased to 268 trillion rupiahs in 2013. Total spending on energy subsidies is always greater than the allocated budget and it often causes problems by the end of each fiscal year. Caraka and Yasin (2014) introduced the government has issued a number of policies to reduce petroleum fuel usage. Crude oil price is based on January 2016 data with 22.48 $/barrel (current price) and it assumed to be rising linearly to 40 $/barrel in the end of 2016. Oil production continues to decline while the demand for energy continues to grow which led to the increase in import of crude oil and petroleum products. This was shown by the deficit 3,5 billion Dollar at oil account in the second quarter which increased from 2,1 billion Dollar deficit in the first quarter of 2014 financial year. On the other hand, fuel subsidy is relatively high, due to increased domestic consumption, the increase in international oil prices and the decline in the exchange rate against the dollar and other foreign currencies. It is estimated that fuel subsidies until the end of 2014 will exceed the budget allocation in 2014. Since the publication of the 2015 edition of the WOO in November last year, the most obvious market development has been the oil price collapse. While the average price of the OPEC Reference Basket (ORB) during the first half of 2014 was over $100/barrel, it dropped to less than $60/b in December 2014 and has averaged close to $53/b in the first nine months of 2015. This new oil price environment has had an impact on both demand and supply prospects in the short- and medium-term, and some lasting effects can be expected in the long-term. Crude oil prices are expected to remain low as supply continues to outpace demand in 2016 and more crude oil is placed into storage. EIA estimates that global oil inventories increased by 1.9 million b/d in 2015, marking the second consecutive year of inventory builds. Inventories are forecast to rise by an additional 0.7 million b/d in 2016 before the global oil market becomes relatively balanced in 2017. The first forecasted draw on global oil inventories is expected in the third quarter of 2017, marking the end of 14 consecutive quarters of inventory builds. In the time domain, the long memory is indicated by the fact that the oil prices eventually exhibit strong positive dependence between distant observations. A shock to the series persists for a long time span even though it eventually dissipates. In the frequency domain, the long memory is indicated by the fact that the spectral density becomes unbounded as the frequency approaches zero.

2. LITERATURE SURVEY

The advent of Globalization has led to a marked increase of the effect prices of goods and services have on one another. Top of the list of goods that have a massive effect on other goods or services is crude oil. Crude oil is arguably the most important commodity traded around the world today. Virtually every sector of the global economy is dependent on crude oil; hence any increase or decrease in the price of crude oil has a ripple effect on the global economy. In view of the importance of crude oil there has been a lot of research and attempts to predict the price of crude oil and her allied products. This is by no means a mean feat as the price of crude oil is non-linear in nature and the task is prone to many difficulties. Tools used by researchers to predict the price of crude oil can be broadly classified into two groups: Using soft computing tools or using econometric tools. In [1] we have an example of the use of soft computing tools particularly neural networks in oil price prediction. The empirical mode decomposition (EMD) technique was used. The EMD uses the Hilbert–Huang transform (HHT) and decomposes a time series to intrinsic mode functions (IMFs). A Feed forward neural network is then used to model the decomposed IMF's and residual components. The results of this neural network are then fed to an adaptive linear neural network (ALNN) which serves as an integrator. The final model is used to forecast both Brent and West Texas Intermediate (WTI) crude oil prices. The correct prediction rate in this work when using the neural network was 69%. Recently in [2] the authors developed an adaptive neurofuzzy interference system that predicts one day ahead whether the price of crude oil (WTI) is going to rise or fall. This hybrid system uses the hybrid learning algorithm and historical data from January 5th,2004 to April 30th, 2007 for training while testing was done throughout the month of May 2007 and also form May to June 2008. The correct prediction rate in this recent work was reported as 66.67%. Support Vector Machines (SVM's) have recently been applied in fields like medicine: breast cancer diagnosis [3], Prediction: wind speed prediction [4], air pollutants level prediction [5] and even the difficult task of oil price prediction [6], [7]. In [6] the authors obtain daily West Texas Intermediate (WTI) oil prices from January 2, 2002 till October, 2008. The data is fed to a Slantlet algorithm which extracts various features as input to the hybrid system consisting of SVM's and Auto Regressive Moving Average (ARMA). In [7] the data used is monthly WTI prices from January 1970 to December 2003(a total of 408 instances) and three systems are designed to predict oil price (Neural Networks, SVM and Auto Regressive Integrated Moving Average: ARIMA). However most of the works on crude oil prediction using support vector machines suffer from a disparity in the training to testing ratio. For instance in [7] the training to testing ratio is an unbalanced 88.2%:11.8%. In [6] the training to testing ratio is 60%:40%. A more appropriate ratio would be closer to 50%:50%. This ensures that the system is not biased. Having reviewed prior works we used training to testing ratio of 50%:50% in our work. Furthermore we used weekly West Texas Intermediate (WTI) spot prices available online on the Energy Information Administration (EIA) website [8].The reason for using weekly as against daily spot prices is to minimize incomplete information due to public holidays or weekends(days when oil is not traded). The paper is organized as follows: in section two we briefly review support vector machines. In section three the oil price dataset is introduced and our novel method to prepare indicators and features from the dataset; to be used as input information for the SVM prediction model. In section four, we describe the SVM model used in this work. In section five, the experimental results of training and testing the SVM model are discussed. Finally, in section six the paper is concluded and areas of improvement for future work are suggested.

3. EXISTING SYSTEM

The grid search is a traditional method for tuning SVR parameters. It has been proved that some meta-heuristic algorithms are more effective than the grid search . It has become a significant trend that meta-heuristic algorithms are introduced into SVR parameter adjustment. Mimicking ethological, biological, or physics phenomena is the main means of meta-heuristic algorithms in solving optimization problems. The algorithms applied to SVR include GA , gray wolf optimization whale optimization algorithm, simulated annealing (SA), etc. The no free lunch theorem in optimization proves that there is and will never be an algorithm to resolve all optimization problems. Hence, brand new algorithms may have the potential to outperform the present ones on some problems.

4. PROPOSED SYSTEM

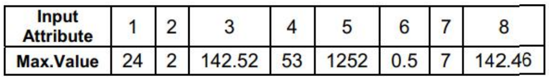
The HGSO approach is a new hand. However, it has showed amazing optimization performance. Compared with PSO, gravitational search algorithm (GSA) , cuckoo search algorithm (CS) , GWO, WOA, EHO and SA, it obtained competitive and superior results . These conclusions provide us with the feasibility and rationality of integratingit into SVR parameter optimization, so that we can combine HGSO and SVR to form a new machine learning approach for prediction. The experimental results also show that proposed approach has excellent comprehensive performance

5. IMPLEMENTATION

A. Data Processing Phase During this phase two important processes take place. Firstly, the values of the eight input features are scaled to values between 0 and 1. Secondly, the output values (predicted oil price) for training and testing are coded to suitable output classes. The scaling technique which is used in this work is based on finding the maximum or the highest value within each input feature for all 1252 observations in the dataset, and dividing all the values within that same feature by the obtained maximum value. Table I shows the maximum values for the input features. In order to create suitable classes for our SVM model we had to code the output. Instead of training the SVM model with output value (weekly oil price), we defined 20 weekly average price ranges within 5 US$/barrel as the output of the SVM model thus giving rise to 20 classes for our model. This is aimed at providing a degree of flexibility to the predicted price and to improve learning of the SVM model. Table II shows the price intervals and their respective output classes. Table III shows examples of the dataset observations prior to scaling; listing the first 10 observations.

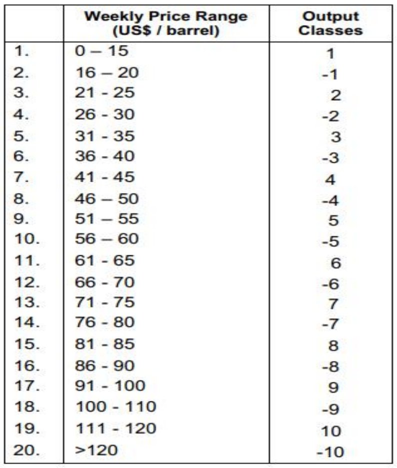
B. SVM Learning Phase During this In the SVM learning phase we used the C-SVM model with an RBF kernel. As stated above the choice of RBF kernel is because it has fewer numerical difficulties, possesses less hyper-parameters than other kernels and is able to handle cases when the relationship between class labels and attributes is highly non-linear as in this case of Oil price prediction. In order to search for suitable parameters (C and) for our RBF kernel we had to perform a parameter search using cross validation specifically the v-fold cross validation method. The cross-validation procedure is a technique used to avoid the over fitting problem. In v-fold cross-validation, we first divide the training set into v subsets all with equal size. Sequentially one subset is tested using the SVM classifier trained on the remaining (v – 1) subsets. Cross-validation accuracy is the percentage of data which are correctly classified. The parameters which produce the best cross validation accuracy are saved and then used to train the SVM learner. The saved model is then used on the out of sample data (testing set). In this work v=5. The parameter search range for C was conducted from (2-50 – 250) while for it was (2-15 – 215). The best C obtained was 2965820 while was 0.00195313. These values of C and were then used for training the SVM learner.

TABLE I. THE HIGHEST OR MAXIMUM VALUE FOR EACH INPUT FEATURE OF THE 1252 OBSERVATIONS; THESE VALUES ARE USED TO NORMALIZE/SCALE THE INPUT DATA PRIOR TO SVM TRAINING



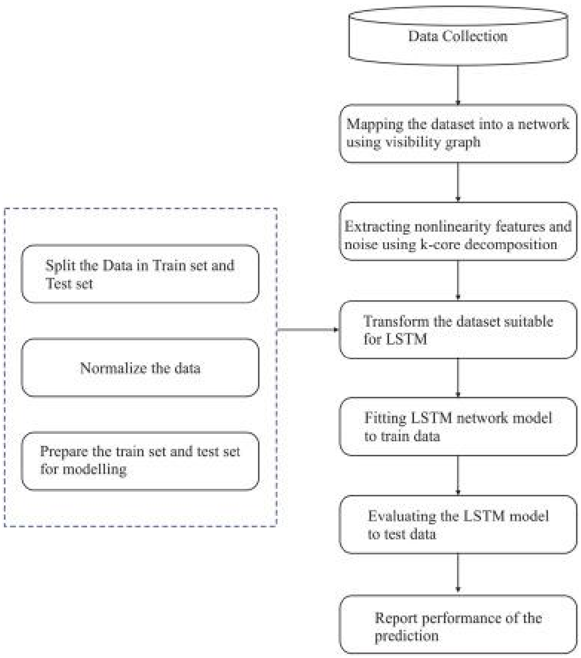
Input Attribute	1	2	3	4	5	6	7	8
Max.Value	24	2	142.52	53	1252	0.5	7	142.46

TABLE II. PREDICTION SYSTEM'S OUTPUT PRICE INTERVAL AND CORRESPONDING OUTPUT CLASSES



	Weekly Price Range (US$ / barrel)	Output Classes
1.	0 – 15	1
2.	16 – 20	-1
3.	21 - 25	2
4.	26 - 30	-2
5.	31 - 35	3
6.	36 - 40	-3
7.	41 - 45	4
8.	46 – 50	-4
9.	51 – 55	5
10.	56 – 60	-5
11.	61 - 65	6
12.	66 - 70	-6
13.	71 - 75	7
14.	76 - 80	-7
15.	81 - 85	8
16.	86 - 90	-8
17.	91 - 100	9
18.	100 - 110	-9
19.	111 - 120	10
20.	>120	-10

The proposed prediction technique comprises of the following steps: mapping the datasets on a network using visibility graph algorithm, extraction of noise from the dataset and determination of the most influential nodes using k-core centrality, finally, LSTM is applied on the extracted datasets to train and test the models. At the end, the prediction of crude oil prices is evaluated with a view to discovering knowledge.


Data Collection
Mapping the dataset into a network using visibility graph
Extracting nonlinearity features and noise using k-core decomposition
Split the Data in Train set and Test set
Normalize the data
Prepare the train set and test set for modelling
Transform the dataset suitable for LSTM
Fitting LSTM network model to train data
Evaluating the LSTM model to test data
Report performance of the prediction

Proposed methodology (prediction technique).

6. EXPERIMENTAL RESULTS

The results of implementing the oil price prediction model were obtained using a 2.2 GHz PC with 2 GB of RAM, Windows XP OS and LIBSVM v 2.9.1. There are a total of 1252 observations which comprises the weekly WTI crude oil prices from January 03 1986 to December 25 2009; available online from the USA- EIA website [8]. We used a training-to-testing ratio of (50%:50%) where the last 626 observations (period: January 02 1998 to December 25 2009) were used for training, while the first 626 observations (period: January 03 1986 to December 26 1997) were used for testing the trained prediction neural model. The reason for using the latest observations or the second period for training rather than testing is explained as follows with the aid of Fig. 1 which shows the graph of the weekly oil prices over the period from 1986 until 2009

TABLE III. EXAMPLES OF PRE- SCALING OBSERVATIONS NUMERICAL VALUES, SHOWING THE INPUT ATTRIBUTES AND CORRESPONDING OUTPUT (OIL PRICE) FOR THE FIRST 10 OBSERVATIONS

Observations	Date	Input Features Value								Weekly Average Price (US$/barrel)
		1	2	3	4	5	6	7	8	
1.	03/01/1986	1	2	0	1	1	0.3	1	26.12	25.78
2.	10/01/1986	1	2	25.78	2	2	0.2	1	26.08	25.99
3.	17/01/1986	1	2	25.99	3	3	0.3	1	24.57	24.57
4.	24/01/1986	1	2	24.57	4	4	0.3	1	20.31	20.31
5.	31/01/1986	1	2	20.31	5	5	0.3	1	19.83	19.69
6.	07/02/1986	1	2	19.69	6	6	0.3	1	16.62	16.72
7.	14/02/1986	1	2	16.72	7	7	0.3	1	16.31	16.25
8.	21/02/1986	1	2	16.25	8	8	0.3	1	14.4	14.39
9.	28/02/1986	1	2	14.39	9	9	0.3	1	14.29	14.25
10.	07/03/1986	1	2	14.25	10	10	0.3	1	12.36	12.27

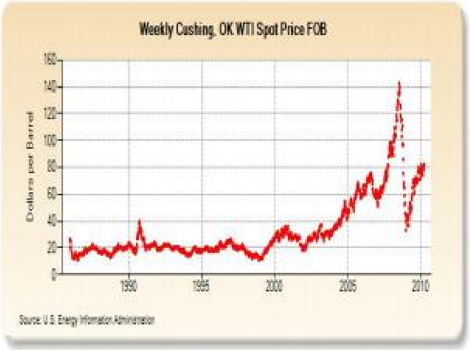
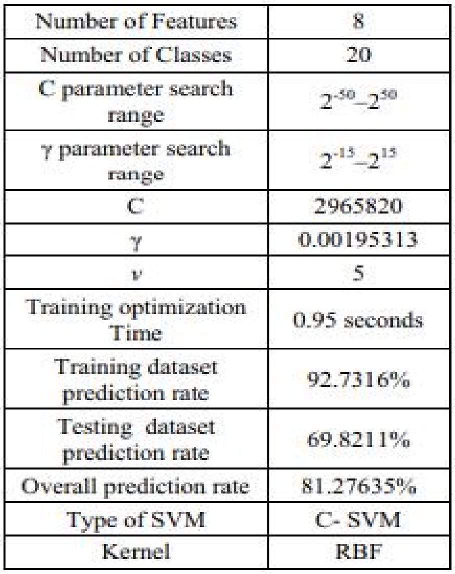


Figure 1. Graph of WTI Spot Prices (January 1986 to December 2009) [8]

Firstly,careful observationofthegraphrevealsthatthe secondperiodismorechaoticandthepricerangeismorevariedthereforetrainingtheSVMmodelwithdatafromthatperiodwouldexpose ittothesevariationsthusmakingitmorerobustwhen usedforprediction.Thefirstperiodpricescanbeseen asmorestableandwithlessrangevariety.Secondly,theoilpricemarketisaneverevolving one,itismorereasonabletotrainSVMmodelwithdatafrommorerecentperiodsinordertoachievemoreaccuratepredictionofthecurrentandfutureprices.TableIVlists thefinalparametersofsuccessfullytrainedSVMmodel,andtheaccuracy rates.Theimplementation resultof trainedpredictionsystemwereasfollows:usingthetrainingdataset(latest626observations)yielded92.7316%accuracyinprediction.

TABLE IV. FINAL PARAMETERS OF THE TRAINED SVM CRUDE OIL PRICE PREDICTION MODEL



Number of Features	8
Number of Classes	20
C parameter search range	2^{-50}–2^{50}
γ parameter search range	2^{-15}–2^{15}
C	2965820
γ	0.00195313
ν	5
Training optimization Time	0.95 seconds
Training dataset prediction rate	92.7316%
Testing dataset prediction rate	69.8211%
Overall prediction rate	81.27635%
Type of SVM	C- SVM
Kernel	RBF

The testing of the trained SVM model using the testing dataset (earlier 626 observations) yielded a correct prediction rate of 69.8211%.

Combining the training and testing prediction results yields an overall correct prediction rate of 81.27635%.

7. CONCLUSION

This paper presented an SVM based prediction system with application to the difficult task of predicting oil prices. This prediction task has been addressed in few previous works that suggested using different econometric models or soft computing methods, with varying degrees of success. In this work, we use WTI average weekly prices of crude oil over the past 24 years as the dataset for developing the prediction system. The output of the trained SVM prediction model provides the weekly average price for crude oil within 5US$/barrel accuracy. The prediction system comprises two phases: Firstly, the data processing phase and the SVM arbitration phase. In the first phase, we apply a novel simple but efficient method of representing data information from the dataset into eight features which we believe have an effect on the oil price. The features include global demand factor and a random world event factor amongst other features. These attributes undergo scaling prior to using them as inputs to the SVM prediction system. We also presented a unique representation to the prediction output using oil prices ranges or intervals of 5 US$/barrel and created corresponding output classes. The second phase is training the SVM classifier. We used a CSVM model with an RBF kernel. The optimal values of parameters C and were searched for and training this model successfully required approximately 0.95 seconds, which is considered as fast. The obtained overall correct prediction rate of 81% is considered as successful taking into account the nonlinearity of the problem. Future work will focus on improving this prediction rate and on comparing the use of SVM to a neural network model when applied to oil price prediction.

8. REFERENCES

[1] L. Yu, S. Wang, and K.K. Lai, “Forecasting Crude Oil Price with an EMD-Based Neural Network Ensemble Learning Paradigm”, Energy Economics vol. 30, no. 5, pp. 2623–2635, 2008.

[2] A. Ghaffari and S. Zare, “A Novel Algorithm for Prediction of Crude Oil Price Variation Based on Soft Computing”, Energy Economics vol. 31, no. 4, pp. 531–536, 2009.

[3] M.F. Akay, "Support Vector machines combined with feature selection for breast cancer diagnosis", Expert Systems with Applications vol. 36, no. 2, pp.3240-3247, 2009.

[4] M.A. Mohandes, T.O. Halawani, S.Rehman and A.A.Hussain,"Support vector machines for wind speed prediction", Renewable Energy, vol. 29, no. 6, pp.939-947, 2004.

[5] W. Wang , C. Men and W. Lu, "Online prediction model based on support vector machines", Neurocomputing vol. 71, no.4,pp.550-558, 2008.

[6] K. He, C. Xie and K.K Lai, "Crude Oil Price Prediction using SlantletDenoising Based Hybrid Method", IEEE International Joint Conference on Computational Sciences and Optimization, 2009.

[7] W.Xie, L.Yu, S.Xu and S.Wang. A New Method for Crude Oil Price Forecasting based on Support Vector Machines, Lecture Notes in Computer Science, vol.3994, pp.444-451, 2006.

[8] Energy Information Adminisntration. Weekly Cushing. OK WTI Spot Price FOB. Retrieved March 14, 2010 from the World Wide

Web:“http://www.eia.doe.gov”.

[9]V.N.Vapnik.Thenatureofstatisticallearningtheory.Statisticsforengineeringandinformationscience.NewYork:Springer,2nded.edition,2000.

[10]C.C.ChangandC.-J.Lin,LIBSVM:alibraryforsupportvectormachines,2001.Software availableat:/http://www.csie.ntu.edu.tw/cjlin/libsvm.

[11]B.N.Huang,C.W.Yang,andM.J.Hwang,“The DynamicsofaNonlinearRelationshipBetweenCrudeOilSpotandFuturesPrices:AMultivariateThres

holdRegressionApproach",EnergyEconomics,vol.31,no.1,pp.91–98,2009.

ABOUT AUTHORS:

KOVVURI LAKSHMI PRIYA DARSHINI is currently pursuing MCA in SVKP & Dr KS Raju Arts & Science College, affiliated to Adikavi Nannaya University, Rajamahendravaram she research interests include Data Structures Web Technologies, Operating Systems, Data Science and Artificial Intelligent.

CH.SRINIVASA RAOis a Research Scholar in the Department of Computer Science & Engineering at Acharya Nagarjuna University, Guntur, A.P, India. He is working as Associate Professor in SVKP & Dr KS Raju Arts & Science College,Penugonda,A.P. He received Master's degree in Computer Applications from Andhra University and Computer Science & Engineering from Jawaharlal Nehru Technological University, Kakinada, India. He Qualified in UGCNET and APSET. His research interests include Data Mining.