LITERATURE SURVEY

In order to become familiar with background, a literature survey was conducted.From this survey,previous work was studied carefully.Number of Research paper are available based on our project.

Authors: Shuang Gao and Yalin Lei proposed the literature on prediction of crude oil price based on stream learning. This paper provides a literature review on the various techniques that have been used to predict crude oil price. They mainly focused on the researches that have used heuristic approach and machine learning techniques in their forecasting study. Therefore, a detailed description of this model was presented in the paper.

The goal of this article is to review the existing literature on crude oil price prediction. They categorized the existing forecasting techniques into econometric models; and then performed an almost comprehensive survey on the available literature with respect to these techniques. A review on the existing literature about crude oil price predictionEconometric models are the most widely used approaches for oil price prediction,which include autoregressive moving average (ARMA) models and vector autoregressive (VAR) models, with possibly different input variables (Pindyck, 1999; Frey et al., 2009).

In this paper, Several machine learning techniques were proposed for oil price prediction, such as artificial neural networks (ANN) (Yu et al., 2008; Kulkarni and Haidar, 2009), and support vector machine (SVM) (Xie et al., 2006). These are nonlinear models which may produce more accurate predictions if the oil price data are strongly nonlinear (Behmiri and Pires Manso, 2013).

However, these machine learning techniques, like other traditional machine learning techniques, rely on a fixed set of training data to train a machine learning model and then apply the model to a test set. Such an approach works well if the training data and the test data are generated from a stationary process, but may not be effective for non-stationary time series data such as oil price data.

In recent years, a new machine learning paradigm called stream learning has emerged to handle real-world applications where 1) there is a continuousflow of data as opposed to a fixed sample of independent and identically distributed (i.i.d.) examples, and 2) the data are generated by a non-stationary process instead of a stationary process (potentially at very high speed). Examples of stream learning applications include social networks, web mining, scientific data,financial data,etc.

The following price forecasting approach have been covered.

Stream learning approach is a supervised machine learning method which uses a set of labeled training data to train an initial model. The main features and advantages of using stream learning for oil price prediction include:

1) The machine learning model will be updated whenever new oil price data are available, so the model continuously evolves over time, and can capture the changing pattern of oil prices.

2) For non-stationary time series data such as oil prices, a forgetting mechanism

(e.g., sliding windows, fading factors) will be deployed when updating the machine learning model.

3) Updating the model requires only a small constant time per new data example, as opposed to re-training the model using the entire training data set.

Stream Approach Procedure:

1. Use the data in the initial training set D to train a machine learning model, denoted by M1, and use M1 to predict the oil price for time slot 1.

2. For time slot t, t = 2, ..., n: Add (xt-1, yt-1) to the training set D and update the machine learning model, denoted by Mt, and use Mt to predict the oil price for time slot t.

MOA (Massive Online Analysis) is an open-source framework software that allows to build and run experiments of machine learning or data mining on evolving data streams (Bifet et al., 2010; Bifet et al., 2012). They use MOA to develop stream learning models for oil price prediction. They have tried several machine learning models, including linear regression, perceptron and regression trees. To tune the parameters of a specific mdel in searching for the optimal solution, they split the initial training data into 90% as the training set and 10% as the development set. The development set is used to tune the model parameters, which is a standard approach in machine learning. They find that peceptron achieves the best accuracy for predicting oil prices and hence they've use it as the core machine learning method within MOA.

Under the traditional machine learning framework, to build a statistically sound machine learning model, the data time series need to be stationary or at least weakly stationary over the whole training and evaluation period. However, the oil price time series are not stationary. Under our stream learning approach, we no longer require that the oil price time series be stationary over the whole training and evaluation period, because the machine learning model will be continuously updated over time to capture the changing pattern of oil price time series. Nevertheless, when we update the machine learning model and apply the model to predict the oil price for the next time slot, there is still an underlying assumption that oil prices are stationary over the relatively shorter time period covering the current training data time slots and the next time slot.

The performance of our streaming learning model and compare it with three popular oil price prediction models.

1. Data and other prediction models

They have use two types of oil price data to evaluate the accuracy of different oil price prediction models. The first one is the U.S. refiner acquisition cost for crude oil imports, which is the weighted average cost of all oil imported into the U.S. It can be viewed as a proxy for world oil price. The second one is the WTI crude oil spot price, which is used
as a benchmark in oil pricing in North America, and is commonly cited in the press. Both data can be obtained from the U.S. Energy Information Administration (EIA) website (EIA, 2016).

2.Performance metrics

They used two standard performance metrics in the oil price predictionliterature for comparing different oil price prediction models. The first metric is Mean Squared Prediction Error (MSPE). MSPE of a prediction model measures the average of the squares of the prediction errors. The prediction error is the difference between the true value and the predicted value.

3.Evaluation

Refiner acquisition cost for crude oil imports. Our stream learning model achieves the lowest MSPE and thehighest DAR among all prediction models for the 1-month, 6-month, 9-month and 12-month forecast time horizons. For example, for the 1-month time horizon, the MSPE of our stream learning model is 11.87, with an error reduction of 31.1% compared with the no-change model (17.23), an error reduction of 13.7% compared with the ANN model (13.75), and an error reduction of 25.3% compared with the forecast combination model in Baumeister and Kilian (2015) (15.89). The DAR of our streaming learning model is 0.594, which is higher than both ANN model (0.582) and forecast combination model (0.570). For the 3-month horizon, although our model has a lower accuracy than the forecast combination model, it is more accurate than the no-change model, and more accurate than the ANN model in terms of MSPE.

References:

- Alavi, A.H., Gandomi, A.H., Lary, D.J., 2016. Progress of machine learning in geosciences: Preface. Geoscience Frontiers 7 (1), 1–2.

- Alquist, R., Kilian, L., Vigfusson, R.J., 2013. Forecasting the price of oil. In: Elliott G,
- Timmermann A, editors. Handbook of economic forecasting. Amsterdam, North Holland, 427–507.
- Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world - a forecast combination approach. Journal of Business and Economic Statistics 33 (3), 338–351.
- Behmiri, N.B., Pires Manso, J.R., 2013. Crude oil price forecasting techniques: a comprehensive review of literature. Alternative Investment Analyst Review 3, 18–29.
- Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B., 2010. MOA: massive online analysis.
- The Journal of Machine Learning Research 11, 1601–1604.
- Bifet, A., Kirkby, R., Kranen, P., Reutemann, P., 2012. Massive online analysis manual.
- EIA, 2016. U.S. Energy Information Administration [Internet]. [cited 2016 Mar 12].
- Nomikos, N.K. and P.K. Pouliasis. (2011). “Forecasting petroleum futures markets volatility: The role of regimes and market conditions.” Energy Economics, Vol. 33, pp. 321–337.