PREDICTION OF FOREIGN EXCHANGE RATE USING REGRESSION TECHNIQUES

Dinesh K. Sharma, University of Maryland Eastern Shore, USA (dksharma@umes.edu)
H.S. Hota, Bilaspur University, India (proffhota@gmail.com)
Richa Handa, Dr. C. V. Raman University, India (richihanda@yahoo.com)

ABSTRACT

This paper explores and compares regression technique with ensemble regression techniques in view of two ensemble learning: Bagging and Boosting (Least Square Boost: LSBoost) for two Foreign Exchange (FX) data namely INR/USD and INR/EUR. The comparative results show that regression ensemble with LSBoost is performing better than others with MAPE =0.6338 for one day ahead prediction at the testing stage. It has also been observed that MAPE value is increasing while increasing value of N in case of N-Days ahead prediction of FX rate.

Keywords: Regression, Foreign Exchange (FX), Ensemble Techniques.

INTRODUCTION

Financial time series data is very chaotic, noisy, fluctuating and nonlinear as different events have occurred in various time periods, so it is very challenging for researchers to develop the accurate predictive model. Prediction for FX rate (Galeshchuk, 2016) is also a very crucial task for N days ahead prediction because of volatile nature of FX Data. Statistical techniques are not able to efficiently predict the FX rate hence different machine learning techniques have been used by many researchers for accurate prediction.

In this research work, we explore two techniques for FX rate prediction: Regression and Ensemble Regression. However, regression technique is widely used for linear data prediction, but not so famous for nonlinear data prediction. On the other hand, ensemble regression technique may improve the performance of predicting model. These two techniques are applied to two different FX data, i.e., INR/USD and INR/EUR to find out N days ahead prediction. Models were ensemble (Opitz et al., 1999) using the ensemble learning techniques: Bagging and Boosting (Freund et al., 1999) and produces more accurate results as compared to the individual model. The original FX rate data has only one feature which may not be able to predict next day FX rate more accurately. Therefore, features are extracted from the original FX rate data using five technical indicators (Handa et al., 2015) suggested by many authors. Experiments were done using K-fold cross validation technique through self-written MATLAB code.

LITERATURE REVIEW

Financial time series data forecasting through machine learning techniques are attracting researchers in the past one decade. Statistical methods, Data mining, and soft computing techniques are the most popular techniques for this. However, recently data mining techniques like decision tree techniques and Artificial Neural Network (ANN) and its variations as soft computing techniques are more powerful techniques than others. These are tabulated in Table 1 as below:

Table 1: Summary of Literature Review

YEAR	REFERENCES	JOURNAL
2017	Hota et al.	International Journal of Computational Intelligence Research
2016	Galeshchuk	Neurocomputing

2015	Simidjievski et al.	Expert system with applications
2015	Handa et al.	International Journal for Research in Applied Science & Engineering Technology
2014	Rehman et al.	ScienceDirect
2011	Pacelli et al.	Journal of Intelligent Learning Systems and Applications
2000	Dietterich	Machine Learning
1999	Schapire	Proceeding of sixteenth International Joint conference on Artificial Intelligence

DATA SET AND METHODOLOGY

Two FX rate data are collected from www.fx.sauder.ubc.ca web source starting from November 9th, 2016 to July 31st, 2017. A total of 189 samples of data set was normalized by dividing each sample with the highest value of the sample in FX data. This preprocessing technique is used to transform the raw data into some meaningful data. Normalization is commonly used preprocessing technique for data smoothing. Equation 1 is used for data normalization, which scales data in the range of [0 1]:

$$X_{new} = \frac{x}{X_{max}} \tag{1}$$

Where x is daily exchange rate of FX data, X_{max} is highest value of FX data and X_{new} is obtained normalize data. In order to generate new relevant and meaning full features, feature extraction technique is applied on normalized FX data based on 5 technical indicators i.e. SMA (Simple Moving Average), EMA (Exponential Moving Average), WMA (Weighted Moving Average), Variance and Standard Deviation as shown in Table 2.

Table 2: List of technical indicators with formulae used to extract features from original feature space.

Technical Indicator	Formulae
Simple Moving Average (SMA)	**SMA**: $\sum_{i=1}^{n} x_i / n$
Exponential Moving Average (EMA)	**Multiplier**: $2/(n+1)$ **EMA**: {Close – EMA (previous day)} x Multiplier + EMA(previous day).
Weighted Moving Average (WMA)	$F_t = \frac{\sum_{i=1}^{n} W_i x_{t-1}}{\sum_{i=1}^{n} W_i}$
Variance	$S^2 = \sum_{i=1}^{n} (x_i - \bar{x}_i)^2 / (n-1)$
Standard Deviation	S= $\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2 / (n-1)}$

Where x_i is actual occurrence for the period of prediction of daily FX rate, n is total number of days in prediction. W_i is the weight to be given to the actual occurrence for the period t-i for WMA calculation. A new data set with new feature space along with 5 extracted features has been prepared where 6 features were used as input and next day FX rate was used as predictor (output). Data is divided in two parts: training and testing. K-fold cross validation technique is used for dynamic data partitioning to make the prediction more accurate.

The following two techniques were used to build learning models:

Regression: Regression is the data mining technique (Han et al., 2012) use to predict numbers only. Regression uses the data where the target value is known. It estimates the relationship between one dependent variable and two or more independent variables.

Ensemble Model: For increasing the prediction accuracy of any predictive model, ensemble models have been developed. An ensemble model is a combination of trained classifiers (Optiz et al., 1999) like the neural network or decision tree. For developing a novel model, the prediction comes from individual classifiers are combined to produce a more accurate result than any single classifier based model. The ensemble is commonly used for machine learning tasks (Simidjievski et al., 2015) like regression and classification. In Regression ensemble two popular methods are used: Bagging and Boosting.

Bagging: Bootstrap Aggregation (Bagging) is used to decrease the variance for increasing accuracy in prediction (Han et al., 2012) using different training sets. In bagging new training data is generated using different training set where some training sets are used more than one times or some training sets may not be included. The bagged model (Breiman, 1996) gives better performance than the single model which produces the final result based on the average predicted value.

Boosting: It is another method of the ensemble. In boosting weight is assigned to each training data (Han et al., 2012). The training tuples may be selected more than one time which is based on its weight. If the tuple is misclassified, then its weight becomes increased and called weak learners, and these weights go for next round to generate training samples to boost the accuracy of the model. If tuples are correctly classified, then their weight becomes decreased. The weights of training tuple decide how accurate tuples are classified. Boosting focus on the weights that are misclassified in the previous round. This is the main idea behind boosting.

RESULT ANALYSIS

Experimental work is done by using a self-written MATLAB generated code for Regression and Ensemble Regression (Bagging and LSBoost). MATLAB function RegressionTree.fit() is used to create Regression tree, Predict() function is used to predict value based on Regression Tree. Fitensemble() method is used for ensemble regression, LSBoost and Bag method is used in the above function as a parameter when Bagging and Boosting methods are used respectively in ensemble regression. Regression and ensemble regression techniques have used for accurate prediction of FX rate with new feature space. FX data obtained after feature extraction were dynamically partitioned into training and testing data sets using k-fold cross validation. Cross validation method (Hota et al., 2017) is better than the static method of partitioning the data. Static partitioning of data with a fixed percentage of training and testing data may bias the learning models on the other hand dynamic partition of data like k-fold partitions data as training and testing dynamically. In k-fold cross validation, the data set is divided into k subsets. In each iteration, one of the k subsets is used as the test set, and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed, and in this manner, each fold takes part in both training as well as in testing. Training and testing data were presented to regression and ensemble regression for N days ahead prediction as 1-Day, 3-Days, and 5-Days. The predicted value of each day is compared against the corresponding actual value of next day FX rate for both FX data: INR/USD and INR/EUR. The accuracy of the predictive model is analyzed by calculating Mean Absolute Percentage Error (MAPE) as shown in Equation 2 and presented in Table 3 and Table 4 respectively for INR/USD and INR/EUR.

$$\mathrm{MAPE} = \frac{\sum_{i=0}^{n} |Y_{a,i} - Y_{p,i}|}{n} \tag{2}$$

Table 3: Comparative results of Regression and ensemble Regression in terms of MAPE for INR/USD at testing stage.

N-Day	Regression	Regression Ensemble

		LSBoost	Bagging
1-Day	0.4294	0.3502	0.4552
3-Day	0.8294	0.6178	0.6205
5-Day	0.8636	0.7159	0.6954

Comparative results of Regression and ensemble Regression in terms of MAPE for INR/EUR at testing stage.			
N-Day	Regression	Ensemble Regression	
		LSBoost	Bagging
1-Day	0.8393	0.6338	0.806
3-Day	1.2391	1.0976	1.0224
5-Day	1.297	1.1937	1.1617

These tables show the results as per expectation, out of two techniques, regression ensemble is producing better results and consistently increasing value of MAPE while increasing the value of N on the other hand out of two ensemble techniques LSBoost is producing better prediction results (MAPE= 0.3502 for 1-Day ahead prediction and MAPE= 0.6178 for 2-Days ahead prediction) while Bagging is producing slightly better results (MAPE= 0.6954 for 5-Days ahead prediction) for INR/USD FX rare data. It has also been observed that there is less fluctuation in INR/EUR FX rate as compared to INR/USD FX rate for the chosen period. The results are also presented in the form of the bar graph as shown in Figures 1.

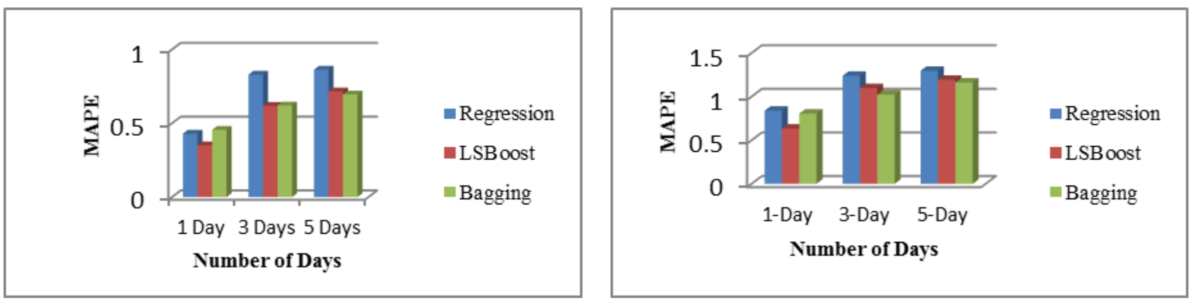

Figure 1: Comparative MAPE of Regression, LSBoost and Bagging for (a) INR/USD FX data and (b) INR/EUR data

CONCLUSION

FX rate prediction using linear model is a little bit tedious due nonlinear nature of data. Regression technique, however, is not so popular, as machine learning techniques especially for time series data prediction which tries to find out best fit linear line with respect to one dependent variable and one or more independent variables. On the other hand ensemble, regression combines more than one techniques together and improves the result in case of nonlinear data up to some extent. This paper explores two techniques for prediction of two FX rate data namely INR/USD and INR/EUR with five extracted features using technical indicators as suggested by authors. Results show that ensemble technique with both bagging and boosting is performing better than regression technique for N-Days ahead prediction. Also, both the ensemble techniques are performing better with slight variations in terms of MAPE.

REFERENCES:

Breiman, L. (1996), Bagging Predictors, *Machine Learning*, Vol. 24, pp. 123-140.

Dietterich, T.G. (2000), An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees:Bagging, Boosting, and Randomization, *Machine Learning*, Vol. 40, 139–157.

Freund, Y. and Schapire, R. E.(1999), A short introduction to Boosting, *Journal of Japanese society for Artificial Intelligence*, Vol. 14,Issue. 5, pp. 771-780.

Galeshchuk,S.(2016), Neural networks performance in exchange rate prediction, *Neurocomputing*, vol.172, pp.446-452.

Handa,R., Hota, H.S. and Tandan, S.R.(2015), Stock Market Prediction with Various Technical Indicators Using Neural Network Techniques, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. Vol.3 Issue. 6, pp. 604-608.

Hota, H.S., Handa, R. and Shrivas, A. K. (2017), Time Series Data Prediction Using Sliding Window Based RBF Neural Network, *International Journal of Computational Intelligence Research*, Vol. 13 Issue 5, pp. 1145-1156.

Han, J., Kamber, M. and Pei, J. (2012), Data Mining Concepts and Techniques, *Morgan Kaufmann*.

Opitz, D. and Maclin, R. (1999), Polpular Ensemble Methods: An Empirical Study, *Journal of Artificial Intelligence Research*, Vol. 11,pp. 169-189.

Pacelli, V., Bevilacqua, V. and Azzollini, M. (2011), An Artificial Neural Network Model to Forecast Exchange Rates, *Journal of Intelligent Learning Systems and Applications*, Vol. 3, pp.57-69.

Rehman, M., Khan, G. M. and Mahmud, S.A. (2014), Foreign Currency Exchange Rates Prediction using CGP and Recurrent Neural Network, *ScienceDirect*, Vol. 10, pp.239-244.

Schapire, R.E. (1999), A Brief Introduction to Boosting, *Proceeding of sixteenth International Joint conference on Artificial Intelligence*. pp. 01-06.

Simidjievski, N., Todorovski, L. and Dzeroski, S.(2015), Predicting long-term population dynamics with bagging and boosting of process-based models, *Expert system with applications*.