The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Financial Market Forecasting by Integrating Wavelet Transform and K-means Clustering with Support Vector Machine

Roshan W. D. S[1], Gopura R. A. R. C[1], Jayasekara A. G. B. P[2] and Bandara D. S. V[1]

[1]Department of Mechanical Engineering, University of Moratuwa, Katubedda 10400, Sri Lanka
sameeraroshanuom@gmail.com, *gopura@mech.mrt.ac.lk*, *sanjaya@mech.mrt.ac.lk*

[2]Department of Electrical Engineering, University of Moratuwa, Katubedda 10400, Sri Lanka
buddhika@elect.mrt.ac.lk

Abstract: Financial market forecasting is a challenging problem and researchers are still exploring the ways to improve the performance of the existing models. This paper presents a forecasting model by integrating wavelet transform, K-means clustering with support vector machine. At the first stage, noise of the input prices is removed by using wavelet denoising. Wavelet multiresolution analysis is used to decompose the original time series in to multiple details and approximated decompositions. Individual support vector models are trained for each detail part. Approximated part is further analyzed by clustering and training support vector models for each cluster. Finally the forecast is made for the wavelet denoised time series by summing up the forecasts of each support vector model. Results have shown that the proposed model has given the accurate forecast and has the capability to support decisions in real world trading.

Keywords: Support vector machine, wavelet transform, K-means clustering, financial market forecasting.

1 INTRODUCTION

Financial market forecasting has been drawn considerable attention in recent years. Accurate forecast of currencies, indices and commodities has become an important issue in decision making in investments. Forecasting financial time series have become a difficult task because of its noisy, nonstationary characteristics and dependence on lot of economic, political environmental and even psychological factors. Forecasters are trying to develop various forecasting techniques in order to reduce the risk of trading. Neural network has become a good alternative over traditional methods like Random Walk model, Box-Jenkins ARIMA, Hidden Markov model etc., because of its capabilities in nonlinear data analysis and universal function approximation Yu et al [1].

Various neural models such as feed-forward, recurrent, radial basis, probabilistic and Support vector machine (SVM) have been developed and researchers are still exploring the ways to improve the forecasting performance of the neural models. Among those models, SVM that is a new kind of kernel based learning machine based on statistical learning theory, has become a good alternative over other neural architectures because of its structural risk minimization property.

In Roshan et al [2], data preparing and architecture determination has identified as the two main problems in building a model for financial market forecasting and new modeling directions have been explored. At the first stage, noise of the signal and the correlation of the inputs should be removed. Thereafter most influential inputs should be selected for the model. Clustering should be done on selected input space where the inputs with similar statistical features are grouped together. Individual models should be created for each cluster and final model should be an ensemble of all individual models.

Therefore this study is focused on improving the forecasting capability of SVM by integrating wavelet transform and k-means clustering according to the above mentioned modeling direction. Section 2 describes the theoretical background of the SVM, K-means clustering and wavelet transform. The simulation results and discussion are given in section 3. Finally the collusions are given.

2 THEORETICAL BACKGROUND

2.1 Support vector machine

The support vector machine (SVM) was proposed by Vapnik [3] and it is based on the structured risk minimization principle. SVM minimize the upper bound of the generalization error by automatically by estimating optimal VC-dimension h with its nested architecture Hayking [4]. SVM is first proposed for pattern classification of two classes and further extended to multi class pattern classification and regression analysis. The SVM regression function can be stated as follows

$$t = W^T \emptyset(X) + b \qquad (1)$$

Where $\emptyset(X)$ is the nonlinear mapping-function. W and b can be estimated as follows.

$$R(C) = C\frac{1}{n}\sum_{i=1}^{N} L_\varepsilon(r_i, t_i) + \frac{1}{2}\|W\|^2 \qquad (2)$$

Where,

$$L_\varepsilon(r_i, t_i) = \begin{cases} |r_i - t_i| - \varepsilon, & |r_i - t_i| \geq \varepsilon \\ 0, & else \end{cases} \qquad (3)$$

r_i is the actual price at the ith period and t_i is the model prediction. $L_\varepsilon(r_i, t_i)$ is called ε-intensive loss function. Term $C\frac{1}{n}\sum_{i=1}^{N} L_\varepsilon(r_i, t_i)$ is the empirical error and the term $\frac{1}{2}\|W\|^2$ is the measure of the smoothness. Tradeoff between empirical risk and smoothness is controlled by C.

Issue to be resolved is to minimize the empirical risk. We can reformulate this optimization problem by introducing two sets of nonnegative slack variables ε_i and ε_i^*. Then the Eq.(2) is can be transformed in to following constrained optimization.

$$\min_{W, b, \varepsilon_i, \varepsilon_i^*} R(W, \varepsilon_i, \varepsilon_i^*) = \frac{1}{2} W^T W + C\{\sum_{i=1}^{N} (\varepsilon_i + \varepsilon_i^*)\} \quad (4)$$

Subject to,

$$W^t \emptyset(X_i) + b - r_i \leq \varepsilon + \varepsilon_i^*$$

$$r_i - W^t \emptyset(X_i) - b \leq \varepsilon + \varepsilon_i$$

$$\varepsilon_i, \varepsilon_i^* \geq 0$$

Taking Lagrangian and conditions for optimality, model can be represented in dual as follows.

$$f(X, a, a^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(X, X_i) + b \quad (5)$$

Where α_i and α_i^* are nonzero Lagrangian multipliers which are the solution the dual problem. $K(X, X_i)$ is the kernel function.

$$K(X, X_i) = exp\left\{-\frac{(X - X_i)^2}{\sigma^2}\right\}$$

Radial basis kernel is utilized in this study with empirically obtained near optimal σ and C

2.2 Wavelet transform

Wavelet transform has proven to be highly effective in analyzing noisy and nonstationary time series by number of researchers. Consider real or complex value continues time function $\psi(t)$ with following characteristics.

1. It is integrable to zero

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0 \quad (6)$$

2. It is square integrable (has finite energy)

$$\int_{-\infty}^{\infty} |\psi(t)|^2\, dt < \infty \quad (7)$$

3. Admissibility condition

$$\int_{-\infty}^{\infty} \frac{|\psi/(w)|^2}{w}\, dw < \infty \quad (8)$$

Where $\psi/$ is fourier transform of ψ

2.2.1 Continuous wavelet transform (CWT)

If a function $\psi(t)$ fulfill above (6), (7), (8) conditions, then the continues wavelet transform can be defined as follows

$$C(a, b) = \int_{-\infty}^{\infty} f(t)\, \Psi_{a,b}(t) dt \quad (9)$$

Where $\Psi_{a,b}(t) = e^{1/2}\, \Psi\left(\frac{t-b}{a}\right)$ is the mother wavelet. a is scale parameter and b is a translation parameter. The relationship between scale and frequency is that low scales correspond to high frequencies and high scales correspond to low frequencies. CWT has much redundant information of the time series and also takes lot of computation time. There for discrete wavelet transform (DWT) has been introduced to eliminate these shortcomings.

2.2.2. Discrete wavelet transform

Definition of the discrete wavelet transform can be given as follows.

$$C(a, b) = C(j, k) = 2^{\frac{j}{2}} \psi_{j,k}(n) \quad (10)$$

Parameters a and b are defined as $a = 2^j, b = 2^j k$

2.2.3 Stationary wavelet transform

Discrete wavelet transform suffers from a drawback of shift variance. That means discrete wavelet transform (DWT) of a signal is differ from the DWT of the shifted version of the same signal due to decimation step involved in the DWT Coifman & Donoho [5]. Stationary wavelet transform is designed by removing decimation step in order to avoid shift variance. It is a necessary property for time series analysis. Therefore Stationary wavelet transform has utilized in this study.

2.3 K-means clustering

In statistics and data mining, K-means clustering is a method of cluster analysis. Importance of the clustering input space is, to remove bias variance dilemma and improve overall generalization. K-means clustering is used to partition n observations into k clusters and each observation belongs to the cluster with the nearest mean. Given a set of observations n with m-dimensional real vector, K-means clustering partitions the n observations into k clusters $(k < n)$ and minimize the sum of squares within-each cluster

$$arg\ min \sum_{i=1}^{k} \sum_{X_j \in S_i} \left\| X_j - \mu_i \right\|^2 \quad (11)$$

where μ_i is the center of the cluster S_i

3 SIMULATION RESULTS

3.1 Forecasting model

In the simulation, model was tested for an actual market condition where new data comes in at every step. In this situation, model has no information what so ever about the next step value. Number of researchers in previous studies has performed financial time series forecasting without taking actual market conditions into consideration. This has led to better testing results but doesn't have capability to make profit in real market conditions

Input selection for financial forecasting often takes a critical place because of the accuracy of the output is entirely depends on the characteristics of inputs. Therefore three types of inputs: time delay price values, simple moving average and exponential moving average, have checked upon the performance of the model. Two moving averages ranges (7 min and 30 min) were tested in the comparison.

Proposed model combines SVM regression, Wavelet decomposition and K-means clustering. Input data is scaled up to reduce the mean squared error of the model. Noise of the input data is further removed by using wavelet denoising, which can be achieved by removing small wavelet coefficients in wavelet decomposition by applying appropriate threshold value, even though averaging methods such as moving averages reduces some amount of noise from the data. Then the approximation and detail parts of the time series have reconstructed separately. Individual SVM models with radial basis kernel and debauchee's wavelet is used for each detailed part.

Radial basis kernel which maps inputs in to an infinite dimensional feature space, has selected according to the Cover's theorem of data mapping Cover [6]. Approximation part is further analyzed by using K-means clustering. Bias variance dilemma is expected to be removed by clustering according to principle of divide and conquer Hayking [7]. Multiple clusters have been created for approximation part of the signal by using K-means clustering. Separate SVM model has created for each cluster. New input vector was assigned to a cluster that has minimum distance from its center and the forecast is made by SVM related to the cluster. Then the output is taken by summing up all SVM models that has been created for detail and approximation parts.

Stationary wavelet transform is highly effective in above steps because of its capabilities in time and frequency localization, orthogonal data mapping and ability to fully reconstruct the original time series. Finally one step ahead and multistep ahead forecast has performed and results were compared with pure SVM regression model and wavelet-SVM regression model.

Pure SVM model has the originally proposed architecture with optimally selected parameters while Wavelet-SVM regression model first decompose original time series in to multiple decompositions and each composition is forecasted using separate SVM regression model and finally results were summed up over all decompositions. Directional statistics (D_stat) and mean squared error (MSE) have used as the performance measures. Model is checked for EUR/USD currency pair which is known to be the most volatile of them all. 1 min values have taken in to the forecast. When removing the noise, 1 min data has the advantage of allowing considerable amount of noise reduction without affecting the profit taking opportunities.

3.2. Results

3.2.1 Pure SVM performance

The results of one step ahead forecast is given in Table 1 and has shown the forecast of pure SVM model is capable of forecasting EUR/USD with great accuracy. But it is only with moving average inputs. Result of the pure SVM is the worst of three models for direct input of time delay data. Even though wavelet denoising has improved the directional performance measure for time series data, it can be clearly seen that wavelet denoising has reduced the accuracy of directional forecast. This is due to the fact that the last few values of the denoised time series are chaotic due to real market conditions. Wavelet transform cannot remove the noise of the values at the edge of the time series. Forecast should be done with those values at every step. Therefore noisy wavelet values at the edge have misled the accuracy of the forecast.

SVM-wavelet model has performed better than pure SVM only when the inputs are past time series data. Again the forecast with moving average inputs is further derogated by wavelet decomposition and wavelet noise reduction. Integration of wavelet transform further increased the error due to the above-mentioned reason.

3.2.2 Importance of the multistep ahead forecast

Even though pure SVM model has shown great accuracy in one step ahead forecasting of moving averages, one step ahead forecasting of moving average data cannot directly used to trade in real markets. It is because of the lag which is an undesirable property of moving average indicators. Therefore one-step forecast accuracy of moving average indicators has no use in real trading systems. However, if the moving average is shifted backwards, the lag will be removed and the cost for removing lag will be a multistep ahead forecast. Chaotic values of the edge of the wavelet coefficients can also be removed by going through same procedure. Again multistep ahead forecast will be needed in order to eliminate the lag. Results of the multistep ahead forecast can be directly used for high frequency trading in real financial markets

3.2.3 Result of multistep ahead forecast.

Results given in the Table 2, shows that the proposed model has outperformed both Pure SVM model and wavelet transformed model in both MSE and D_stat in multistep ahead forecast. Table 2 shows the results for 5 step ahead forecast which was the best performed among multistep forecast. Directional accuracy and MSE of the pure SVM has greatly reduced in the multi-step ahead forecast. Elimination of chaotic values of the wavelet decomposition leads to better directional accuracy of SVM-wavelet model with time series data. Even though 30 min moving average has led to better performance of the model, large period of averaging may has the disadvantage of missing sharp peaks. Thus selection of moving average should be optimized along with the profit of the system.

4 CONCLUSION

This study propose a reliable financial forecasting model by integrating wavelet transform k-means clustering and support vector machine. Even though wavelet transform isn't suitable for single step ahead forecast, it can be highly effective with multistep ahead forecast. Proposed model has the ability to forecast multi step ahead values in with great accuracy. Simple moving average with wavelet noise reduction has over performed other two types of input

Table 1. Single step ahead forecast

Model	Time series data		Simple Moving average		Exponential moving average	
	Direct	Wavelet denoising	Direct	Davelet denoising	Direct	Wavelet denoising
Pure SVM	MSE:1.2278*10e-7 D_stat:42%	MSE:1.2463*10e-7 D_stat:48%	Range:7 MSE:8.3206*10e-9 D_stat:79% Range:30 MSE:8.0944*10e-9 D_stat: 90%	Range:7 MSE:9.3827*10e-9 D_stat:77% Range:30 MSE:8.5179*10e-9 D_stat:87%	Range:7 MSE:9.2117*10e-9 D_stat:71% Range:30 MSE:9 .0204*10e-9 D_stat: 82%	Range:7 MSE:9.0384*10e-9 D_stat:68% Range:30 MSE:7.7411*10e-9 D_stat:84%
SVM-Wavelet model	MSE:3.335*10e-7 D_stat:51%	MSE:5.4623*10e-8 D_stat:64%	Range:7 MSE:2.076*10e-8 D_stat:76% Range:30 MSE:8.2192*10e-9 D_stat:88%	Range:7 MSE:2.3938*10e-8 D_stat:77% Range:30 MSE:1.4058*10e-8 D_stat:88%	Range:7 MSE:1.3290*10e-8 D_stat:70% Range:30 MSE:3.8440*10e-9 D_stat:81%	Range:7 MSE:1.2873*10e-9 D_stat:66% Range:30 MSE:2.14331*10e-9 D_stat:82%
Proposed model	MSE:1.032*10e-7 D_stat:54%	MSE:1.3623*10e-8 D_stat:71%	Range:7 MSE:2.1569*10e-9 D_stat:79% Range:30 MSE:1.2653*10e-9 D_stat:89%	Range:7 MSE:2.5692*10e-9 D_stat:78% Range:30 MSE:1.3506*10e-9 D_stat:89%	Range:7 MSE:2.3526*10e-9 D_stat:71% Range:30 MSE:1.255*10e-10 D_stat:82%	Range:7 MSE:1.9348*10e-10 D_stat:70% Range:30 MSE:0.8047*10e-10 D_stat:83%

Table 2. Multi step ahead forecast (5 steps)

Model	Time series data		Simple Moving average		Exponential moving average	
	Direct	Wavelet denoising	Direct	Wavelet denoising	Direct	Wavelet denoising
Pure SVM	MSE: 1.2658*10e-6 D_stat:44%	MSE: 9.5684*10e-7 D_stat:60%	Range:7 MSE: 1.5437*10e-6 D_stat:65% Range:30 MSE: 1.3268*10e-6 D_stat: 79%	Range:7 MSE: 9.3087*10e-7 D_stat:72% Range:30 MSE: 9.0026*10e-7 D_stat:82%	Range:7 MSE: 1.6789*10e-6 D_stat:63% Range:30 MSE: 1.4869*10e-6 D_stat: 77%	Range:7 MSE: 9.7831*10e-7 D_stat:72% Range:30 MSE: 9.0865*10e-7 D_stat:80%
SVM-wavelet model	MSE: 2.5684*10e-6 D_stat:49%	MSE: 8.4568*10e-7 D_stat:78%	Range:7 MSE: 2.4561*10e-7 D_stat:80% Range:30 MSE:8.2192*10e-8 D_stat:85%	Range:7 MSE: 1.2986*10e-7 D_stat:81% Range:30 MSE: 2.5687*10e-8 D_stat:87%	Range:7 MSE: 2.3658*10e-7 D_stat:79% Range:30 MSE:3.5842 *10e-8 D_stat:84%	Range:7 MSE: 1.96952*10e-7 D_stat:81% Range:30 MSE: 1.8965*10e-8 D_stat:85%
Proposed model	MSE:5.6938 *10e-7 D_stat:52%	MSE: 2.3568*10e-8 D_stat:85%	Range:7 MSE: 5.5147*10e-8 D_stat:81% Range:30 MSE: 9.956*10e-9 D_stat:91%	Range:7 MSE: 2.3614*10e-8 D_stat:85% Range:30 MSE: 8.7695*10e-9 D_stat:93%	Range:7 MSE:6.5387 *10e-8 D_stat:80% Range:30 MSE: 1.3568*10e-9 D_stat:87%	Range:7 MSE:3.068 *10e-8 D_stat:84% Range:30 MSE: 1.069*10e-9 D_stat:89%

based forecasts. Model is tested by simulating real market conditions. Hence it is suitable for decision making in a real trading system. Performance of the model can be further improved by optimizing noise reduction level and threshold, number of clusters, multi-step size, moving average span and SVM parameters.

REFERENCES

[1] Yu, L., Wang, S. & Lai, K.K., 2005. Adaptive smooth Neural Network in foreign exchange rate forecasting. *LNCS*, p.523 – 530.

[2] Roshan, W.D.S., Gopura, R.A.R.C. & Jayasekara, A.G.B.P., 2011. Financial forecasting based on artificial neural networks: Promising directions for modeling. *6th IEEE International Conference on Industrial and Information Systems (ICIIS)*, pp.322 - 327.

[3] Vapnik, V.N., 1995. *The nature of statistical learning theory*. USA: Springer.

[4] Hayking, S., 1999. Committee Machines. In *Neural Networks: A comprehensive foundation, 2nd ed.* NJ, USA: Prentice Hall. Ch. 2.

[5] Coifman, R. & Donoho, D.L., 1995. Translation invariant de-noising. *Lecture Notes in Statistics*, p.125–150.

[6] Cover, T.M., 1965. Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14, pp.326-34.

[7] Hayking, S., 1999. Learning processes. In *Neural Networks: A comprehensive foundation, 2nd ed.* NJ, USA: Prentice Hall. Ch. 2. pp.97-98.