World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:5, 2012

Building a Trend Based Segmentation Method with SVR Model for Stock Turning Detection

Jheng-Long Wu, Pei-Chann Chang, and Yi-Fang Pan

Open Science Index, Computer and Information Engineering Vol:6, No:5, 2012 publications.waset.org/4222.pdf

***Abstract*—**This research focus on developing a new segmentation method for improving forecasting model which is call trend based segmentation method (TBSM). Generally, the piece-wise linear representation (PLR) can finds some of pair of trading points is well for time series data, but in the complicated stock environment it is not well for stock forecasting because of the stock has more trends of trading. If we consider the trends of trading in stock price for the trading signal which it will improve the precision of forecasting model. Therefore, a TBSM with SVR model used to detect the trading points for various stocks of Taiwanese and America under different trend tendencies. The experimental results show our trading system is more profitable and can be implemented in real time of stock market

***Keywords*—**Trend based segmentation method, support vector machine, turning detection, stock forecasting.

I. INTRODUCTION

ONE of the important issues for forecasting market trend is to detect a turning point when the stock prices go through the up/down cycle. The turning point can be applied to make trading decisions in stock investment. In the real world, the turning point detection is very complex since there are lots of factors affecting the movement of the stock price. The stock price variations are even under a high level of noise influence. These factors include interest rates, economic environment and so on. Owing to the development in Computational Intelligence tools, a collaborative model is developing to predict or detect the turning points based on these factors and thus making profitable trading for investors.

In recent years many researches consistently achieve returns since they used forecasting approach in Computational Intelligence tools to reduce investment risk under historic information of stock market [1] [2]. In the financial market technical index have been applied to explain the stock price variation and even used to detect turning point for stock trading such as William Index or Relative Strength Index… etc. In traditional forecasting, financial researches usually use mathematic model to predict stock trend or price, but the stock price was quickly changed in a very short time therefore these models cannot immediate adjusts to these dynamic changes. In the last few years, several representations of time series data have been proposed, the most often used representation is piecewise linear representation (PLR) [3] [4] [5]. It used to decompose a historic stock price data into a series of bottom and peak points is good [1] [2]. But the PLR has a problem that it is not consider the multiple trends of trading in past literatures, if we want to improve the forecasting model for fit to the real environment which the segmentation method need to detect multiple trends [5].

Additionally, the trading model development in machine learning (ML) has provided more approaches for predication such as artificial neural network, fuzzy system, support vector machine and other Artificial Intelligence (AI) approaches. From the previous researches we find that the support vector machine (SVMs) outperforms than other MLs approaches in a lot of researches. Also ML has shown better performances than others as in application areas such as [1] [2] [6]. The support vector regression (SVR) is a regression-based model based on support vector machine. SVR model has high toleration error rate and high accuracy for learning solution knowledge in complex problems [6]. Therefore we want to build a new segmentation method with SVR for a stock trading system. The main purposes of this research are:

1) Developing a trend based segmentation method (TBSM) for improving SVR forecasting model to easily capture the trading knowledge in stock market. In the new segmentation algorithm we have been considered multiple trends such as up-trend/down-trend and hold-trend to find turning points in stock price.

2) Providing investor simple trading decision according to our trading system of TBSM with SVR to generate trading decisions, i.e., buy/sell/hold under daily basis.

II. LITERATURE REVIEW

In past years most of the studies shows that financial time series forecasting usually focus on fundamental or technical analysis. Fundamental analysis could forecast the long term of trade in stock market which may only trade once a year. Therefore, the technical analysis is applied to further improvement in the long term analysis. Technical indices could provide the trade information in short or medium term. Stock technical indices are very commonly applied in financial analysis. An intelligence stock trading system [7] is developed which according to these Technical indices and probabilistic reasoning to forecast trading points. The conventional approach to modeling stock market forecasting is using the univariate time series and they include multiple regression, autoregressive (AR), moving average (MA), GARCH and ARIMA [8] models.

Jheng-Long Wu is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (corresponding author, phone: 886-3-4638800 Ext.2768; fax: 886-3-4352077; e-mail: jlwu.yzu@gmail.com).

Pei-Chann Chang is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: iepchang@saturn.yzu.edu.tw).

Yi-Fang Pan is with the Department of Information Management, Yuan Ze University, Taiwan, R.O.C. (e-mail: s996235@mail.yzu.edu.tw).

Open Science Index, Computer and Information Engineering Vol:6, No:5, 2012 publications.waset.org/4222.pdf

A strong trading signal detection method must consider not only stock price variations, but also other information that can help investor. As this section describes, these models can help to solve the trading signal problem. However, these models require complex mathematical formulas which are not easy to be understand by the investors. Therefore, it seems there should be another way to solve this problem more efficiently. The stock market forecasting problem is very complex, in recent years the financial time series forecasting appears more new approaches which are computer-based to learning knowledge in pattern recognition task.

A. Support Vector Regression

Machine learning techniques have been applied for assigning trading signal. Many studies used support vector machine for determining whether a case contain particular class [1] [2] [9] [10]. But the shortcoming is only deal with discrete class labels, whereas trading signal is continuum because a weight of signal can take a buy or sell power. Grounded in Statistical Learning Theory [11], support vector regression is capable to predict the continuous trading signal while still benefiting from the robustness of SVM. SVM have been successfully employed to solve forecasting problems in many fields, such as financial time series forecasting [12], emotion computation [13] and so on. There are still some significant parameters (C, ε, σ) in empirical results and the value is given by experiment in this research. SVM apply in financial time series which is focused only on price or trend to forecasting that cannot direct to transaction in stock market. Therefore, we build a signal for transaction which call trading signal that the value range from zero to one for decision making

B. Trend Based Segmentation Method

In the time series database there are many approaches such as Fourier transform, Wavelets, and piecewise linear representation to find the turning point on time series data. PLR used to support more tasks and provide an efficient and effective solution [14]. The major functions that using some of piecewise to represent various trends. PLR has been applied to stock market for stock price forecasting, the result shows used PLR could improve forecast accuracy [15]. Additionally, the PLR is the good segmentation approach to found stock price trend but in the turning points problem it is not easy represent the multiple trends for the trading strategy. Zhang and Siekmann [16] presented an approach for segmentation method based on multiple trends to detect stock data. Therefore, in this research, we will using multiple trends segmentation method to segment the trading pairs which has three trends as represent the real trend of stock price. From this concept, we will capture the good trading signal for take a high profit in the stock market.

III. A TBSM WITH SVR FORECASTING MODEL FOR STOCK MARKET

This paper attempts to build forecasting model by a TBSM with SVR model which follow the supervisor learning procedure development. The framework of trading model has five stages: the first is generating trading segments by TBSM approach from historical data; the second is trading signal transformation from trading segments; the third is feature selection by SRA approach; the fourth is learning the forecasting model by SVR approach. The trading point detection from SVR predicating data, show in Fig. 1.

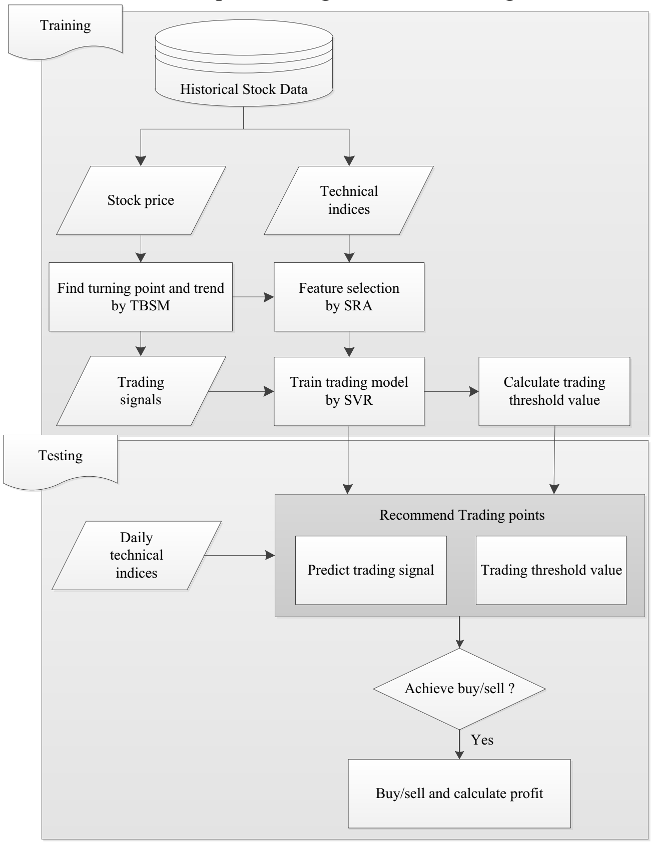

Fig. 1 The framework is TBSM with SVR for stock trading decision

The trading system detail processes as follows:

A. Generating the Trading points by TBSM

One of the main tasks of the present research is to define a proper similarity measure between subjective sequential segment data. According to the characteristics of sequential data, a piecewise linear representation of the data is appropriate. A variety of algorithms to obtain a proper linear representation of segment data have been presented [14] [17]. However, this stage will find the good trading pair from the stock price for the learning of forecasting model as SVR. In this work there is a special considerations must be taken into account when selecting the cut point where a linear model will be fitted over the data. From the pseudo-code of Fig. 2 we can know the steps that it considers three trends including up-trend, down-trend and hold-trend to form a fitted segment.

Therefore, segmentation results from TBSM that it has a lot of trading pairs from the historical stock price. These segments are belonging to three trends. By the way, the range of stock price of each segments are different, so we need to transform the segments to the consist range of trading signal.

Open Science Index, Computer and Information Engineering Vol:6, No:5, 2012 publications.waset.org/4222.pdf

```
Define :  "Threshold"  is cutting threshold
          "X_Thld"  is horizontal area
          "Y_Thld"  is vertical area

Procedure TBSM(T)
Let T be represented as X[1,2…n], Y[1,2…n]
n = 0
Draw a line between (X_1, Y_1) and (X_n, Y_n)
Max d = maximum distance of (X_i, Y_i) to the line
If (Max d > Threshold)
   Let (X_i, Y_i) be the point with maximum distance
   For j = X_1 : X_n
      If ( |X_j − X_i| < X_Thld ) and ( |Y_j − Y_i| < Y_Thld )
         Then Point[n] = [X_j, Y_j], n = n + 1
      End If
   End For
   Select from Point[n] : X_s1 = Min(X_n) , X_s2 = Max(X_n)
   Return :  S1 = T[X_1, X_s1]
             S2 = T[X_s2, X_n]
End If
```

Fig. 2 A new Trend based segmentation method (TBSM) for time series data

B. Trading Signal Transformation

In this step, the aim is calculating the trading signal in a time series of a stock. From last step we can give a lot of segments of trends. We suppose a segment of S_k is up-trend then the range of value of is change to [0,0.1,…,1], if S_k is hold-trend then the range of value of S_k is change to 0.5, if S_k is down-trend then the range of value is change to [1,0.9,…,0]. IF S_1, S_2 and S_3 are up-trend, hold-trend and down-trend, respectively, then like to $S = \{S_1, S_2, S_3\}=\{<0,0.5,1>,<0.5,0.5>,<1,0.5,0>\}$. Now we transform each time series segment S_k to trading signal segment $S_k^{'}$. Final we combine these $S_k^{'}$ to a fully time series of trading signal. If the segment belongs to up-trend or down-trend then using the formula of Eq. 1 transform to trading signal.

$$x_i = \begin{cases} i/L & \text{if } S_k \text{ is uptrend segment} \\ (L-i)/L & \text{if } S_k \text{ is downtrend segment} \end{cases} \quad (1)$$

Where L denotes the length of segment S_k. whereas segment S_k belong to hold-trend is using the formula of Eq. 2.

$$x_i = \begin{cases} 1 & \text{if } i\text{th is higher point in time series} \\ 0 & \text{if } i\text{th is lower point in time series} \\ 0.5 & \text{otherwise} \end{cases} \quad (2)$$

The result of trading signal for S shows in Fig.3. The red dotted line is the Hold-trend which is a special signal for increase reflects on the original turning points, so the Hold-trend is not a horizontal line. The purple dotted line is down-trend signal and the orange dotted line is up-trend signal. For example, in the time series T the T_1 to T_5 and T_{10} to T_{14} are Hold-trend signal representation, T_6 to T_9 is down-trend signal representation, and finally T_{15} to T_{18} is up-trend signal representation.

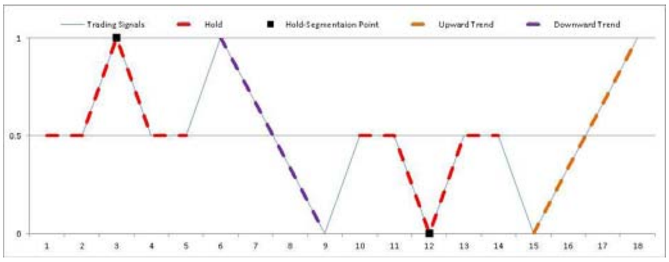

Fig. 3 An example of trading signal of a time series

C. Feature Selection by SRA form Technical Indices

According to reference shows if select some of important feature could improve effect before learning model [18]. Therefore, we want to use stepwise regression analysis to select features from technical indices that selected features have high relational feature with trading signal. In addition, the range of the input variables of TS and SVR model should be between 0 and 1. Hence, the selected technical indices are normalized as follows:

$$Normal(x_{ij}) = \frac{x_{ij} - Min(x_i)}{Max(x_i) - Min(x_i)} \quad (3)$$

$i = 1,\ldots,n;\ j = 1,\ldots,m;\ n \text{ and } m \in \Re$

Where, $Normal(x_{ij})$ denotes the normalized value of j^{th} data point of i^{th} technical index. $Max(x_i)$ denotes the maximum value of i^{th} technical index. $Min(x_i)$denotes the minimum value of i^{th} technical index. x_{ij} denotes original value of j^{th} data point of i^{th} technical index. n and m denotes the total number of technical indices and data points respectively.

In the feature selection part input factors will be further selected using step-wise regression analysis (SRA). The SRA has been applied to determine the set of independent variables which is most closely affecting the dependent variable. The SRA is step by step to select factor into regression model which if factor has the significance level then it is selected. We can follow Eq. (4) to calculate the F value of SRA.

$$F_j^* = \frac{MSR(x_j \mid x_i)}{MSE(x_j \mid x_i)} = \frac{SSR(x_j \mid x_i)}{SSE/(n-2)(x_j \mid x_i)} \quad i \in I \quad (4)$$

$$SSR = \sum(\hat{Y} - \overline{Y})^2 \quad (5)$$

$$SSE = \sum(\hat{Y}_i - Y_i)^2 \quad (6)$$

where SSR denotes a regression sum of square. SSE denotes residual sum of squares. x is the value of technical index. y is the value of stock price. n is the total number of training data. $\hat{Y}$ denotes the forecasting value of regression. $\overline{Y}$ is the average stock price of training data. After the feature selection by SRA, we can provide a set of features to form an input vector for the next step to learning the forecasting model.

D. Learning the Forecasting Model by SVR Approach

To learn the function that maps the feature values to the predicted trading signal, we use SVR with ε-k insensitive loss function where an errors small than a pre-defined parameter is considered. We use a nonlinear kernel to learn a nonlinear forecasting model because of the stock complex stock will be

Open Science Index, Computer and Information Engineering Vol:6, No:5, 2012 publications.waset.org/4222.pdf

better. The input vector from technical indices after feature selection of SRA and output vector form trading signal by TBSM with transformation.

E. Detecting Trading points from Forecasted Signal

In the daily forecasting, if the forecasted trading signals by SVR satisfied a buy threshold then this means needed to buy stock quick because of very close turning point otherwise if the state satisfied a sell threshold then need to sell stock. These satisfied points are recommending to transaction in stock market. We will calculate the buy/sell threshold values for two trading types. The trading thresholds of two types follow as:

$$Buy\_threshold = \mu + \sigma \tag{7}$$

$$Sell\_threshold = 1 - \mu + \sigma \tag{8}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i^{'} \tag{9}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i^{'} - \mu)} \tag{10}$$

Where μ denotes the average of trading signal in training data. σ denotes the standard deviation of trading signal in training data. *Buy_threshold* denotes the buy trading threshold. *Sell_threshold* denotes the sell trading threshold. If forecasted trading signals form SVR model in testing data is more than *buy_threshold* then this is a suggest trading point for buy stock, else if forecasting signal in testing data is smaller than *sell_threshold* then this is a suggest trading for sell stock.

From the Fig. 4 we show that how to suggest the buy/sell points for one stock in a time series which is the red square point is the buy points, green triangle point are the sell points. Both are satisfied two thresholds which the orange dotted line is sell threshold and the purple dotted line is buy threshold, so we can daily forecasting the trading point in an automatically trading system by this propose model.

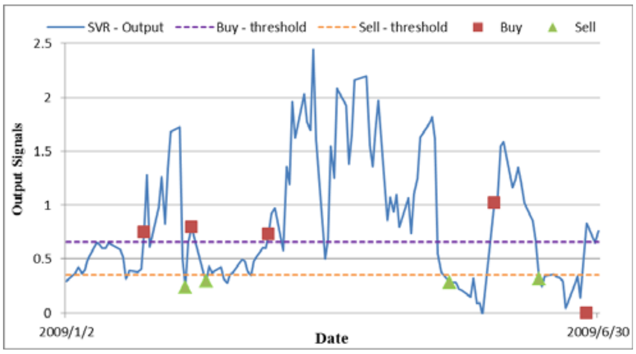

Fig. 4 An example of result for detecting trading points

IV. Experimental Results

In the exprimental result setion, we have selected 7 stocks for analysis which is collected from American stock market that data period of stock price and technical indices collected were from 1/2/2008 (dd/mm/yy) to 6/30/2009, which uses the period of 1/2/2009 to 6/30/2008 to form the testing data period. Due to there is no any standard method to calculate the model effect or efficient in stock market when predict the buy-sell signal. In this paper, we calculate the profit how much profit we can capture. The profit is calculated follow as Equation (11).

$$profits = M\prod_{i=1}^{k}\left\{\frac{(1-a-b)\times C_{S_i} - (1+a)\times C_{B_i}}{(1+a)\times C_{B_i}}\right\} \tag{11}$$

where M is the total amount of the money to be invested at the beginning, a refers to the tax rate of i^{th} transaction, b refers to the handling charge of i^{th} transaction, k is the total number of transaction, C_{S_i} is the selling price of the i^{th} transaction and C_{B_i} is the buying price of the i^{th} transaction.

The TABLE I shows the profit of three models in six stocks. Our propose model as TBMS-SVR has high profits better than other forecasting model, the highest profit is 92.34% in Apple stock and lowest profit is 13.95% in JNJ stock. From the overall the best average profit is TBSM-SVR, therefore our propose TBSM-SVR trading system is a best stock turning point forecasting model.

TABLE I
Compare to Rate of Profit Between Multiple Stock Trade Models

Stock	TBSM-SVR (RBF kernel)	BRR-ES	PLR-BPN
Apple	**92.35%**	61.28 %	12.97%
BA	**59.49%**	38.08 %	17.50%
CAT	**43%**	-23.83 %	9.36%
JNJ	13.95%	-16.86 %	**16.88%**
S&P 500	**22.78%**	-4.56 %	3.77%
VZ	**28.6** %	15.36 %	27.72%
Xom	**22.4%**	0%	-1.99%
Avg.	**40.37%**	9.92%	12.32%
St.D	**0.28%**	0.31%	0.1%

From the Fig. 5 there have four buy/sell trading pairs, in which three trading pairs have excess rate of return, and one trading pair not. According to the trading result on Apple stock we can earn 92.35% profit by our proposed trading system by TBSM with SVR.

Fig. 5 A time series of stock price of Apple stock for buy/sell trading points

V. Conclusion

In this paper we proposed trading sytem of TBSM with SVR that has been effectively captured the high profit and stability. Experimental results shown the TSBM has segmented trading trends is very well in stock market. If we considerd the multiple trends in segment approach then we can more closed to read trends. Additionally it has proved a good trading signal for SVR model to learning trading knowledge. However, the primary goal of the investor could be easily achieved by providing him with simple trading decisions.

REFERENCES

[1] P.C. Chang, C.Y. Tsai, C.H. Huang, and C.Y. Fan, "Application of a case base reasoning based support vector machine for financial time series data forecasting, " in International Conference on Intelligent Computing, South Korea, 2009, pp. 294-304.

[2] P.C. Chang, C.Y. Fan, and C.H. Liu, "Integrating a piecewise linear representation method and a neural network model for stock trading points prediction," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 39, 2009, pp. 80-92.

[3] X.P. Ge, "Pattern Matching in Financial Time Series Data," Computer Communications, 27, 1998, pp. 935-945.

[4] E. Keogh and M. Pazzani, "An enhanced representation of time series which allows fast and accurate classification," clustering and relevance feedback. In: KDD, 2009, pp. 239-243.

[5] V. Lavrenko, M. Schmill, D. Lawrie, P.Ogilvie, D. Jensen, and J. Allan, "Mining of concurrent text and time series," In: Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining, 2000, pp. 37-44.

[6] Alex J. Smola and Bernhard Schölkopf, "A tutorial on support vector regression. Statistics and Computing," vol. 14, 2004, PP. 199-222.

[7] Depei Baoi and Zehong Yang, "Intelligent stock trading system by turning point confirming and probabilistic reasoning," Expert System with Applications, vol. 34, 2008, pp. 620-627.

[8] Allan Timmermann and Clive W. J. Granger, "Efficient market hypothesis and forecasting," International Journal of Forecasting, vol. 20, 2004, pp. 15-27.

[9] Y.W. Wang, P.C. Chang, C.Y. Fan, and C.H. Huang, "Database classification by integrating a case-based reasoning and support vector machine for induction," Journal of Circuits, Systems, and Computers, vol. 19, no. 1, 2010, pp. 31-44.

[10] Li Zhang, W. D. Zhou, and P.C. Chang, "Generalized nonlinear discriminant analysis and its small sample size problems," Neurocomputing, vol. 74, no.4, 2011, pp. 568-574.

[11] V. Vapnik, "The nature of statistical learning theory," Springer, 1995.

[12] Nicholas I. Sapankevych and Ravi Sankar, "Time series prediction using support vector machines: a survey," IEEE Computational Intelligence Magazine, vol. 4, 2009, pp. 24-38.

[13] J.L. Wu, L.C. Yu, and P.C. Chang, "Emotion classification by removal of the overlap from incremental association language features," Journal of the Chinese Institute of Engineers 34(7), pp. 947-955, October, 2011.

[14] Y. Zhu, D. Wu, S. Li, "A piecewise linear representation method of time series based on feature pints," The proceeding of KES 2007/ WIRN 2007, Part II, Vietri sul Mare, Italy, pp. 1066- 1072, January , 2007.

[15] Huanmei Wu, Betty Salzberg, and Donghui Zhang, "Online event-driven subsequence matching over financial data streams," The proceedings of the 2004 ACM SIGMOD, 2004, pp.23-34.

[16] Z. Zhang and J. Siekmann, "Pattern Recognition in Stock Data Based on a New Segmentation Algorithm," Springer, 4798, 2007, pp. 520–525.

[17] J. Lachaud, A. Vialard, F. de Vieilleville, "Analysis and comparative evaluation of discrete tangent estimators," In: Andres, E., Damiand, G., Lienhardt, P. (eds) DGCI 2005. LNCS, Springer, vol. 3429, 2005, pp. 240-251.

[18] Thomas J. Burkholder and Richard L. Lieber, "Stepwise regression is an alternative to splines for fitting noisy data," Journal of Biomechanics, vol. 29, no. 2, 1996, pp.235-238.

Open Science Index, Computer and Information Engineering Vol:6, No:5, 2012 publications.waset.org/4222.pdf