2015 IEEE 14th International Conference on Machine Learning and Applications

The Influence of Sample Reconstruction on Stock Trend Prediction via NARX Neural Network

Yi Wei
Department of Computer Science and Engineering
University at Buffalo
Buffalo, NY, USA 14260
yiwei@buffalo.edu

Vipin Chaudhary
Department of Computer Science and Engineering
University at Buffalo
Buffalo, NY, USA 14260
vipin@buffalo.edu

Abstract—in our study, through the established NARX neural network model, the sample data of stock AAPL in NASDAQ from 2006/01/01 to 2015/01/01 are utilized for training. The results show that under the same sampling frequency, with the increase of MA period, the trend of volatility becomes lower with obvious longer time delay, which will help to predict the trend of movement. In addition, through the use of reconstructed data containing the trend information as training sample, it has significantly reduced the prediction error, which is 16.29% lower than using daily training sample and 16.90% lower than using weekly training sample. The outputs directly reflect the probability of trend movement at every time point in stock price. It also improves the generalization ability of NARX model, so as to predict the stock trend change at a certain time. It has successfully estimated the possibility of buying and selling points, which provides the necessary theoretical basis on how to determine the stock trading points.

***Keywords—trend prediction*; *stock trading points*; *NARX*; *sample reconstruction*; *artificial neural network*.**

I. Introduction

The stock market is characterized by both high risk and high return simultaneously. The financial industry has been pursuing the goal of characterizing this for decades. But the stock market is a complex system affected by a plethora of diversified uncertainties, whereas its volatility of price reflects strong stochasticity and randomness. The stock market can be influenced by multiple variables and dependencies and its trend has high degree of uncertainty. Thus, the derived indices and indicators may imply the intrinsic movement that helps determine the stock value.

The artificial neural networks (ANN) simulate the human brain by a series of interconnected memory and topological elements and have been widely used to model complex non-linear systems. In recent years, stock investors have taken advantage of the development of ANN for quantitative, scientific analysis and equity investments evaluation. Specifically, ANNs have the following characteristics that make them particularly suitable for stock trading:

(1) Distributed information storage; Parallel data processing.

(2) Self-learning; Self-adaptation.

(3) Implicit expression of nonlinear relationship.

Finance researchers are taking a keen interest in the application of neural networks. Bankruptcy prediction, debt-risk assessment and security market applications are three areas that are heavily researched in the finance industry [1]. Many studies have shown that, through the study on historical data, neural networks have the ability to find the relevant dependency of the market movement. For example, the neural network model can be used for prediction of the currency exchange rate, such as Dollars, Japanese Yen, German Mark, British Pound, Swiss Franc and the Australian Dollar [2]. In addition, some researchers utilized the time delay neural network (TDNN), recurrent neural network (RNN) and probabilistic neural network (PNN) for stock market trend prediction [3]. Of course, selecting a proper network model is the key factor in predicting investment probabilities [4]. In addition, the effective result of data analysis can be necessarily guaranteed by selecting proper financial technical indicators [5].

The stock market usually shows intense volatilities and hidden dependencies as well as randomness and unusual behavior to some extent. Therefore it is necessary to establish a neural network model not only for the evaluation on the current movement of stock market, but also for the trend prediction of the development of the stock market in future.

II. Methodology

2.1 Neural Network Model

Neural network has the features in terms of self-adaptive, self-organizing, associative memory, fault tolerance and robustness, etc., without the need to establish a specific mathematical model, and without considering the factors regarding the sequence complexity and non-linear features. Obviously, the neural network is a powerful tool for non-linear prediction. Each time-series sequence can be used as an independent non-linear sequence with its corresponding input-output system. E. Diaconescu illustrates that the NARX

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.210

(non-linear autoregressive models with exogenous input) is suitable for processing chaotic time series [6]. It has been reported that gradient descent learning can be more effective in NARX networks than in other recurrent architectures with hidden states, it also has the strong power on intensive computing[7]. Unlike conventional recurrent networks, NARX is a neural network model which is computationally equivalent to Turing Machine and without any computational loss [8].

2.2 Design of Neural Network

In order to find the reliable intrinsic feature between neural network and time-series data in multi-period, the selected sample data trained in its neural network should be inevitably improved according to the characteristics of the model of time-series sequence. In this paper, the NARX neural network is utilized to predict the time-series data. There are various approaches on the implementation of the estimation of the function F_{NARX} for NARX model. The model in our study is a Feed-forward Neural Network with Tapped Delay Line (TDL) including an input layer with two TDLs and hidden layer with ten neurons (Fig. 1).

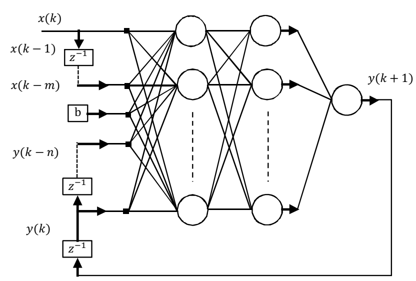

Fig. 1. NARX model with tapped delay line and recurrent network

NARX is an important class of discrete nonlinear system, where $F_{NARX}(\cdot)$ is a nonlinear function, which can be expressed as followed:

$$y(n+1) = F_{NARX}\begin{pmatrix} x(n), x(n-1), \dots, x(n-d_x+1); \\ y(n), \dots, y(n-d_y+1) \end{pmatrix} \quad (1)$$

Where $x(\cdot) \in R, y(\cdot) \in R$, represent the input and output of the model in discrete-time form, in order to maintain its generalization, the dead-time parameters is set to 0. Thus, the input vector is defined as [x (n); y (n)].

2.3 Stock Sample Data

In our research, the stock AAPL from 2006/01/01 to 2015/01/01 including 2265 trading day has been analyzed, which was collected from Yahoo Finance Website http://finance.yahoo.com/. The selected in-market trading data is composed of daily closing price. Some factors in terms of ex-right, ex-dividend, etc., are not taken into consideration. For the stock AAPL data, 70% is used for training, 15% is used for validation, and 15% is used for test. The above data analyzed for NARX prediction in Matlab are split into three data sets including training set with 1585 samples, validation set with 340 samples and test set with 340 samples.

The stock closing price is treated as a discrete time-series sequence. Let P_k be the stock price on k-day, then:

$$P_k = f(P^k) = f(X_{k-1}, \dots, X_{k-t}) \quad (2)$$

Where $P_k = (P_{k-1}, \dots, P_{k-t})^T$, $f(\cdot)$ is a non-linear continuous function. $f(\cdot)$ describes the dynamics mechanism of creating a time-series sequence. Therefore, we are able to make predictions on the time-series sequence once a proper $f(\cdot)$ can be found. However, it is very difficult to capture the particular dynamic process due to the complexity of the stock market movement.

2.4 Sample Data Reconstruction

The traditional technology and approaches for the prediction of time-series data requires the selected test data to be stationary, but most of the time-series stock data is non-stationary. For such a non-stationary stock data sequence, to improve the situation during training phase, it is effective to obtain the changes in the short-term trend in the sequence at a certain time involving moving average (MA). In financial markets, the moving average is an important indicator that reflects the trend and movement of stocks, etc., which plays an important role in data analysis that cannot be underestimated [9]. In the NARX model, we will use the pre-processed training samples as input; utilize the mean of adjacent data in the sequence to indirectly reflect the trend stock movement. Our implementation is to calculate the recent local mean at time t in the time-series data so that it reflects the trend of movement between the value of the sample at time t and its average within $t-N$ period.

In the time series sequence, let the moving average of closing price data in the sampling frequency $f = \frac{1}{N}$ be as followed:

$$MA_N = \frac{\sum_{i=1}^{N} P_i^N}{N} \quad (3)$$

Define the trend changes between the moving averages in sampling frequency $f_1 = \frac{1}{N}$ and $f_2 = \frac{1}{M}$ as followed:

$$F_{\mathrm{Trend}}(P_i) = \left(f_1 \cdot \sum_{i=1}^{N} P_i^N - f_2 \cdot \sum_{i=1}^{M} P_i^M \right) = \begin{cases} +1, \Delta > 0 \\ -1, \Delta < 0 \\ 0, \Delta = 0 \end{cases} \quad (4)$$

In formula 4, P_i represents the closing price in each moment and describes the relationship of moving averages of training samples with period T in different sampling frequency. It can be seen that between different period T, the training sample exists without the trend of movement, but in order to train the NARX model, it is necessary to pre-process the sequence, because the sample data having different sampling frequencies cannot map to each other correspondingly in time. For the sample with sampling frequency f_1 and f_2, when $f_1 < f_2$, at time t_0, sample f_1 has corresponding data, while sample f_2 does not exist. In other words, two such sample sequences at time t_0 are not comparable to each other. The sequence mapping in time is implemented such that the sample data with larger sampling frequency should be mapped to the sample data with smaller sampling frequency.

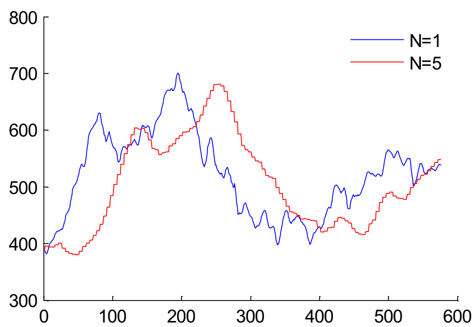

Fig.2. The data trend of movement for stock AAPL in different sampling frequencies.

Note: x-axis represents the trading days, y-axis represents the MA of stock price. Blue line represents the 5-day MA, red line represents 5-week MA.

If the data in weekly sequence has the same time span as the daily sample sequence, it will be extended in time so as to map the daily sample (Fig.2). Samples having different sampling frequency in the given time period will have different reactions to the change of trend. To extract the trend change, using sample data with certain frequency, the variation among MA in different period of sequence at time t can be estimated, and can be extracted at every moment containing trend information. As the target set for training, according to formula 4, let +1 represent the state of buying point, which means the positive signed variation at time t in the time-series sequence; let -1 represent the state of selling point which means the negative signed variation at time t in the time-series sequence; let 0 represent the state of non-existence of trading points, which means no distinct variation captured at time t in the time-series sequence.

III. Learning Process

3.1 Training Algorithm in Different Sampling Frequency

In order to meet the training requirements, in this study, we are using the dynamic back-propagation algorithm. Some researchers believe that the dynamic back-propagation algorithm is required to compute the gradients, which is more computationally intensive than static algorithm and takes more time. Furthermore, the error surface of dynamic networks can be more complex than those for static networks. Training is more likely to be trapped in local minima. However, the training method for NARX, takes the advantage of availability at the time of output values, which is crucial for prediction of time-series sequences [10]. According to the above-mentioned features, we have developed the following training algorithm:

Double-sampling training algorithm
(1) Initialize input $U(x)$ and target $T(y)$ sequence.
(2) Select samples at frequency f_1 randomly as input sequence.
(3) Calculate the input and output of the hidden layer.
(4) Calculate the input and output of the output layer.
(5) Calculate the generalization error of the output layer.
(6) Calculate the generalization error of the hidden layer.
(7) Update weights
(8) Randomly select samples for training and return to step 3, until finished
(9) Randomly select samples at frequency f_2 $(f_1 < f_2)$ and return to step 3, until the global error is minimum

The above training algorithm is adaptive to the data with different sampling frequency that is for training purpose. Whenever output is obtained at time t, it will be trained again as input along with external input at time $t+1$ in the sequence.

3.2 Training Inputs

In the recursive network, it will have a greater impact on the prediction in the near future when the training output from each step contains trend information. The target data is defined as the trend category $\langle +1|-1|0\rangle$ corresponding to its closing price which has the trend probability of increase, decrease and stay. According to the category, the trading points are determined by the underlying changes of trend at the particular moment in sequence. Two sets of data in our study – the closing price with sampling frequency in single day and multiple days are used for model training. The difference of the two data sets is that, within the same time period, they carry different number of samples in sequence, respectively, as the training input. The training data sequence is converted to vector form establishing a row matrix with eigenvalues. Select two data sequence in N period and M period.

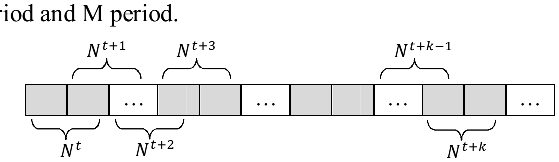

Fig.3. N-day period sampling

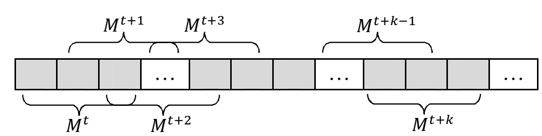

Fig.4. M-day period sampling

The first data sequence is defined as N day moving average (Fig. 3) and the second data sequence is defined as M day moving average (Fig. 4).

3.3 Training Outputs

The training process on time-series data in neural network model, relies on the stationary property of data, however, the stock data does not have this feature. Having the local trend of movement extracted from training sample, we can prepare the target data produced from the corresponding changes in the trend. The previous defined category values of +1, -1 and 0 represent the upper and lower limit of the trend, where +1 is the 100% probability of going up, -1 is the 100% probability of going down, and 0 is the unknown state or 100% probability of stay. Therefore, using the characteristic of intrinsic value of stock data analysis, we use the target sequence to describe the range of stock trend movement, which is defined as followed:

$$SEQ_{target} = \bigcup_{i=1}^{N} \langle +1|0|-1 \rangle_i \qquad (5)$$

The target data is a discrete sequence composed of +1, -1 and 0, with its value reflecting the probability of trend movement. For such target data, it clearly records trend information (formula 5). We use the output sequence to describe the distribution of probability of stock trend movement, which is defined as followed:

$$SEQ_{output} = \bigcup_{i=1}^{N} F_{NARX}(P_i \mid F_{Trend}(P_i)) \qquad (6)$$

Where P_i denotes the input value and $F_{Trend}(P_i)$ denotes the upper and lower limits of trend of moving averages. In the well-trained network model based on a given target, the predicted output describes the probability distribution, for any new coming input; its probability of trend ranges from -1 to +1 (Fig.5).

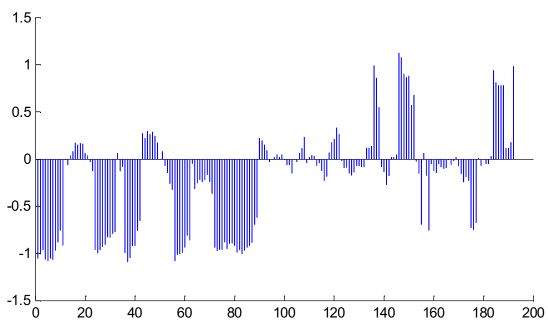

Fig.5. Probability distribution of stock trend movement

3.4 Generalization Ability Improvement

Although NARX had been shown to be universal approximators, it was found that NARX had difficulty modeling seasonal patterns in time series [8]. This is the also the problem with using the neural network to model stock data. During stock prediction, it is possible that there are many outliers existing in the data, which make it difficult to train the network. Therefore, when a time series contains significant seasonality, the data need to be de-seasonalized. Considering outliers influence evaluation, it is difficult to see which part of data makes more contribution to the prediction result. Thus, that is determined through actual experiments.

IV. Results

4.1 The Influence of Sample on Prediction

In general, the time series data are analyzed using mathematical methods, utilizing models such as AR, MA, and ARMA and usually obtaining satisfactory performance on predictions. However, it is difficult to establish an ideal system model because of uncertainty and chaos in nonlinear characteristics of time series in practical applications. In addition, even if the training data have better performance on the network model, it is possible that the trend of the time series in the future may significantly change, thus, the prediction will produce higher deviation.

In order to obtain short-term trend of time series at a particular time point, adapting to the non-stationary characteristic, we introduced the moving average process on sample data, to help extract the trend features. From the result of analysis, through the MA processing, NARX has reduced the prediction error.

4.2 The Influence of Multi-Sampling on Prediction

Due to the continuity of information in time series, in order to see macroscopic trends, we need to sample data at fixed intervals but varying the sampling window. Setting appropriate sampling frequency will help find a steady increasing or decreasing local trend, which leads to correct trend prediction. It also helps to determine the specific stock trading points. To extract trend information from the time series of different frequencies, the moving average with fixed period is applied in the sequence at different frequencies. Then calculate the moving average in different periods and compare the following:

1. The difference of MA samples with the same frequency (daily) in different periods (5days, 10days).
2. The difference of MA samples with different frequency (daily, weekly) in the same period (5days).

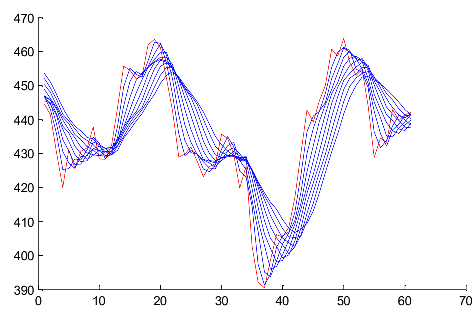

Fig.6.The trend movement of MA in multiple periods

Note: x-axis represents the trading days, y-axis represents stock price

The results show that under the same sampling frequency, with the increase of MA period, the change in trend becomes slower and the trend of latency becomes more obvious (Fig. 6), but such a delay helps to capture the change in trend. Fig. 7 shows that the training error gradually reduces with the increase of MA periods and that larger period of MA will provide less trends and characteristics accordingly.

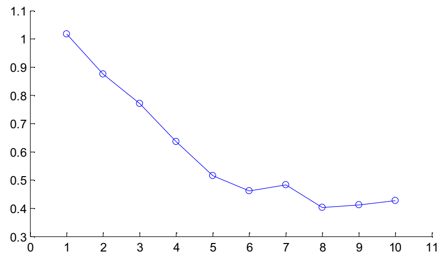

Fig.7. The relationship between MA periods and training errors

Note: x-axis represents the trading days, y-axis represents stock closing price

4.3 The Influence of Trend Extraction to Training

Trend extraction is the key step in time-series data analysis. J. Wijsen believes that the trend is a pattern of moving direction, and proposed the concept of trend dependency [11]. Considering the trend of dependency will help in obtaining more intuitive information from time series data.

Training individual stock price without involving the period in different frequency as time series data in analysis makes it difficult to accurately predict its future trend. This is because the selected data do not contain enough features that reflect the impact of price movement. In this study, we mark the trend movement features from the difference between samples having different frequencies as signed unit value, utilizing trend extraction on the processed sample data.

TABLE 1. THE COMPARISON OF PREDICTION ERROR ON NARX

	AAPL Prediction MSE		
# of Training	Daily Sample	Weekly Sample	Reconstructed Sample
Mean	0.271	0.277	0.108
1	0.205	0.327	0.125
2	0.248	0.266	0.095
3	0.209	0.261	0.093
4	0.305	0.286	0.090
5	0.331	0.251	0.116
6	0.325	0.284	0.102
7	0.274	0.269	0.127
8	0.245	0.262	0.122
9	0.336	0.233	0.087
10	0.231	0.331	0.121

In order to capture the potential movements, the selected samples in 1-5 day pairs and 1-5 week pairs are constructed to keep consistency with daily MA and weekly MA. The lower MSE shows good generalization ability (Table I), which is 16.29% lower than that of daily training sample and 16.90% lower than that of weekly training sample.

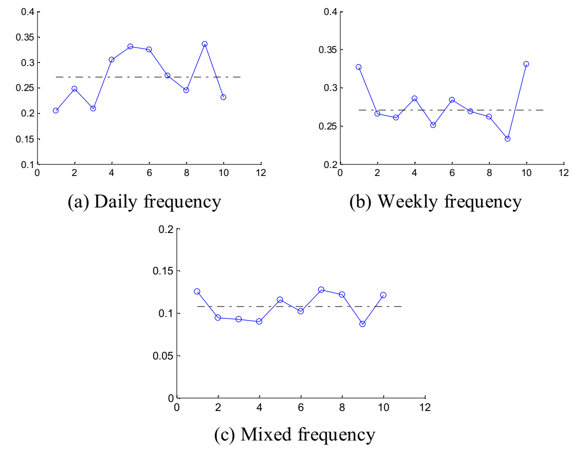

Fig.8 The prediction MSE of stock AAPL data in different frequency and period

Note: x-axis represents the number of training, y-axis represents the MSE

4.4 Probability Distribution of Trend Prediction

The similarity matching of time series is the basis of time-series pattern matching problem; its difficulty of implementation lies in the definition of measuring similarity, and the computing speed of matching algorithm. If we apply pattern-matching algorithms directly on the entire time series, the computational performance is poor. To improve search speed, feature extraction can be used on the time series. E. Keogh has proposed a piecewise linear approximation approach on the time series; it uses piecewise curve fitting to extract linear features in time series [12]. This method can effectively pre-process the time series, which is widely used in trend extraction. In this study, we use signed marks as the target data to describe the change direction of stock prices (Fig. 9).

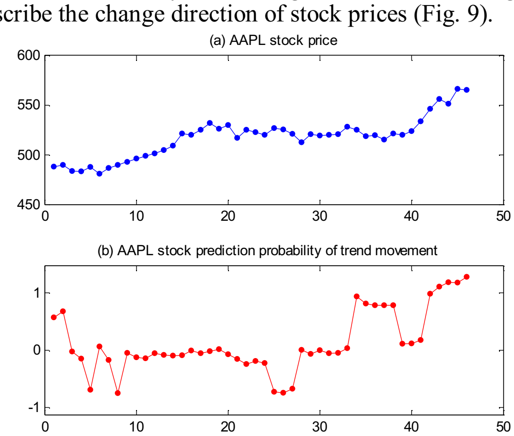

Fig.9 The trend movement of stock prices for stock AAPL

Note: x-axis in both graphs represents the trading days, y-axis in upper figure represents the real stock price, and y-axis in lower figure represents the probability of trend movement.

The process of data reconstruction is to analyze and extract trend information from input and target data. It is necessary to make assumptions on the trend movement and reconstruct sample data for the model to generate the prediction series of movement probability. Using such reconstructed target data for training, the predicted outputs – movement probability will fall between the intervals$[-1, +1]$. The data reconstruction helps the model to generate the detailed movement probability, which is the probability of future trend changing at a certain time instance (9 (b)), comparing to the actual price of stock AAPL (9 (a)), the outputs directly reflect the probability of trend movement at every time point in stock price.

V. Conclusions

Stock trend prediction is difficult due to huge data and complex relationships of internal inference, thereby requiring complex training algorithms. Although many problems have been solved using neural networks, they lack a systematic approach for stock trading.

To reduce the training error, we have proposed a method by using multi-frequency sample, through the use of trend information at different frequencies. Since the sampling frequency has a great impact on stock prediction, it is better to configure the sampling frequency to find the stable local trend, so that it helps to determine stock trading points. Our results show that under the same sampling frequency, with the increase of MA period, the trend of volatility becomes lower with longer time delay, which helps predict the trend of movement. The samples with different sampling frequencies are quite different in training error. Multi-frequency samples of moving average produce low training error, which compensates the trend information to single frequency sample data.

In the established NARX model, we have used the moving averages that can reflect the important trend of the stock in the training data as input, to effectively extract the underlying trend in time series, so as to take advantage of the average of the adjacent data to indirectly reflect the stock trends. Training model using reconstructed samples with trend information helps to reduce prediction error, which is 16.29% lower than using daily training sample and 16.90% lower than using weekly training sample.

The reason why NARX has good performance on time series data prediction is that the trend extraction has played an important role. Since the sample data do not contain enough features of price fluctuations, using such data will produce inaccurate result without the use of trend extraction. In this study, the labeling process on the difference of data in different sampling frequencies trains the features in the trend, which is to extract the trend from training data. The results show that the prediction error can be significantly reduced and the prediction accuracy can be improved if using the data with trend information in NARX model. Through our proposed NARX neural network model, we have trained the sample data of stock AAPL in NASDAQ from 2006/01/01 to 2015/01/01. The results show that, if we use the data with the trend information as training set, it will significantly reduce the prediction error and improve the generalization ability of NARX model, so as to predict the stock trend change at a certain time.

In addition, the results of training the nonlinear time series data show that the outputs in probability series form with respect to original price has practical significance. During stock trading, it is sometimes difficult for investors to make decisions, because its price is affected by a variety of factors. Therefore, providing the analytical results in an expected probability form to investors is one of the main goals for researchers. Through the training on nonlinear time series data, the network model has provided the probability sequence corresponding to its price movements, which is an important guidance for stock trading. It has successfully estimated the possibility of buying and selling, which provides the necessary theoretical basis on how to determine the stock trading points.

Reference

[1] T.S. Quah and B. Srinivasan. Improving returns on stock investment through neural network selection. Expert Systems with Applications, 17: 295-301, 1999.

[2] J. Yao and C.L. Tan. A case study on using neural networks to perform technical forecasting of forex. Neurocomputing, 34: 79-98, 2000.

[3] E.W. Saad, D.V. Prokhorov and D.C. Wunsch. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. Neural Networks, IEEE Transactions on, 9(6): 1456-1470, 1998.

[4] T. Kišasondi and A. Lovrenčić. Selecting neural network architecture for investment profitability predictions. Journal of Information and Organizational Sciences, 30: 1, 2006.

[5] A. Wysocki and M. Ławryńczuk. An investment strategy for the stock exchange using neural networks. Computer Science and Informat, 183-190, 2013.

[6] E. Diaconescu. The use of NARX Neural Networks to predict chaotic time series. WSEAS Trans. Comput. Res. 3(3): 182-191, 2008.

[7] B.G. Horne and C.L. Giles. An experimental comparison of recurrent neural networks. In Advances in Neural Information Processing Systems 7, G. Tesauro, D. Touretzky, and T. Leen, Eds. Cambridge, MA: MIT Press, 697-704, 1995.

[8] H.T. Siegelmann, B.G. Horne and C.L.Giles. Computational capabilities of recurrent NARX neural networks. IEEE Transactions on Systems, Man and Cybernetics, Part B, 27(2): 208-215, 1997.

[9] Y. Wei and V. Chaudhary. Fast quantitative analysis of stock trading points in dual period of DMAC. IEEE Conference on Big Data Computing Service and Applications, 34-43, 2015.

[10] D.P. Mandic and J.A. Chambers. Recurrent neural networks for prediction. John Wiley & Sons, 304, 2001.

[11] J. Wijsen. Trends in databases: reasoning and mining[J]. IEEE Transactions on Knowledge and Data Engineering, 13(3): 426-438, 2001.

[12] E. Keogh. A fast and robust method for pattern matching in time series databases[C]. Proceedings of the 9th International Conference on Tools with Artificial Intelligence. Newport Beach: IEEE, 578-584, 1997.