Intraday Stock Forecasting

Victor Louwerse, Léon Rothkrantz

Abstract: *The objective of this study is the development of a forecasting system for intraday stock price movement. Here, intraday refers to stock movement within a single trading day. To identify trends and forecasting market movements, artificial neural networks (ANN) are employed by the forecasting system. In a simulation study we compared the performance of our system with a buy & hold and naive strategy. The model has been tested in some experiments on real data and the results are reported in this paper.*

Key words: *Stock Market, Intraday, Time Series Forecasting, Neural Network.*

INTRODUCTION

The stock market, a phenomenon dating back to over 200 years with the creation of the New York Stock Exchange in 1792, has ever since been about trading, investing, making and losing money. Over the years, investment banks, regular banks, private investors and trading houses have specialized themselves in the art, science and skill of sensible and profitable investing and trading on the stock markets. With the advancement of computer technology, researchers have been utilizing the increased computing capacity to assist stock trading. Stock forecasting is an area of much interest in machine learning.

The core objective is to create an intraday stock forecasting system. As will become evident most existing research does not focus on intraday forecasting, but on day-to-day forecasts instead. Intraday forecasting is a major challenge. There exists limited research in computer-based intraday stock forecasting. Existing studies use very different models or radically closer forecasting horizons. A similar challenge arises with the employment of technical analysis, since virtually all previous research uses technical indicators on a day-to-day basis, while the proposed models applies them on an intraday level.

The relevance of the stock forecasting problem is ever present in financial business. Some of them merely manage stock portfolios, while others make countless trades per day as their primary means to profit. Such high-volume trading corporations rely on their financial experts and automated systems to perform their trades. They make use of algorithmic trading, which is automated trading based on a set of simple rules. An intraday stock forecasting system deployed in such an environment has the potential to increase profit, by allowing more direction in trading strategies, moving beyond simple rules.

A reliable/accurate intraday stock forecasting system with a decision support system on top of it could simply assist financial experts in making their trading choices, or it could be attached to automated systems trading businesses already have in place. The proposed variable forecasting horizon would make the system more flexible in its applications, providing either a general guide for the direction of trades (further forecasting horizon), or a guide for immediate buy and sell decisions (closer forecasting horizon).

The outline of the paper is as follows. In the next section the state-of-the-art in time series and stock forecasting is presented. Then our model for intraday stock forecasting is presented and its implementation. The model has been tested in some experiments and the results will be reported. The paper ends with an evaluation and final conclusion.

LITERATURE STUDY

This section explores the state-of-the-art in the field of time series forecasting with an emphasis on stock forecasting. The paper of Zhang et al. [1] is a classic in ANN-based forecasting and one of few thorough survey papers in its field. Even though it may appear dated, most of its discussions are still very relevant. The paper starts off with a general introduction to ANNs in forecasting, followed by an outline of the history of ANN in forecasting, from the first application in 1964, to the introduction of the back propagation algorithm in 1986, to more recent work often comparing ANN performance with that of statistical and regression-based models. Thereafter, the actual forecasting capabilities of neural networks with respect to other types of models are detailed, based on the findings of surveyed papers. Considered methods include Box-Jenkins and ARIMA models. The conclusions of surveyed papers with regards to the effectiveness of ANN vary. The study notes that inferior cases of ANN performance may often be due to incorrect application / procedure. In conclusion, the paper states that the unique properties of neural networks make them suitable for forecasting problems, even though it is inconclusive whether they outperform classic methods. Inconsistencies in literature dealing with ANN are ascribed to the trial-and-error approach typically employed by researchers. Also, future research directions regarding more structural methodology are recommended.

In [2] Adya et al. researched the effectiveness of Neural Networks at forecasting and prediction. This paper reviews 48 studies concerning ANN-based business-related forecasting. The study concludes that ANNs show potential for forecasting and prediction, if effectively validated and implemented. 19 of the 22 studies (86%) that were well-validated and either well-implemented or having some problems with implementation, showed ANN to outperform alternative methods. The authors express some reservations towards this result, since studies with negative or mixed results with implementation problems were excluded. Furthermore, they recommend future research to spend more efforts towards validity, as over half the initially considered studies were insufficiently validated.

The extensive survey of De Gooijer et al. [3] regards general time series forecasting, during the past 25 years. It encompasses 380 papers and 20 books. The major techniques that are covered are exponential smoothing, ARIMA-models, non-linear models (including ANN), long memory models and ARCH/GARCH models, which are used mostly for volatility forecasting. Furthermore, the paper reflects on forecast evaluation and accuracy measures, techniques for combining different forecasting methods, prediction intervals and densities.

The study of Lee described in [4] details the iJADE Stock Advisor. Before it gets to the details of the actual stock forecasting, it provides an extensive outline of the iJADE framework and its web-mining capabilities. The stock forecasting itself is facilitated by a hybrid RBF recurrent neural network, which was previously developed by the same authors. It is an extension of the conventional RBF network, adding two main features: First, it introduces a 'forgetting factor' in the back propagation algorithm. Secondly,a 'decay' is introduced in the recurrent time difference mechanism. Both additions effectively increase the importance of more recent observations.

Paper [5] of Tsang et al. presents a straightforward ANN-based stock forecasting solution, starting off with an introductory literature review. Concepts of fundamental versus technical analysis and the efficient market hypothesis are visited. Traditional time series forecasting using linear models is discussed, as well as the recent popularity of ANNs in stock movement forecasting. Included is an outline of the basic principles behind ANNs.

MODEL

This section provides first the theoretical background for the forecasting models developed in this paper. The central component in each forecasting model is an artificial neural network with a fully connected feed-forward architecture. The forecasting models differ in their training methods. This section clarifies the use of technical indicators as input data and the corresponding buying value as target output, which is a summary of future stock performance.

Fundamental analysis (FA) aims to determine the intrinsic value (actual economic worth) of commodities by analysis of the underlying economic variables, which drive market movement. These vary from macro-economic variables, such as national unemployment or interest rates, to micro-economic variables, such as a corporation's quarterly profits. Ultimately, FA will determine if a commodity is currently under-priced or over-priced. The fundamental analyst expects the commodity's market price to move towards its intrinsic value. Under this expectation, a commodity should be bought when under-priced and sold when over-priced.

While fundamental analysis studies the causes of market movement, technical analysis (TA) studies its effects. Technical analysis is the study of market action for the purpose of forecasting future price trends, where market action refers to properties such as price and volume. The primary tools of TA are technical indicators and charts, which must be interpreted to obtain buy and sell signals. Technical indicators are often simple, sometimes complex formulas using stock price, volume or other data as inputs, and providing a value or sequence of values as output. They can be applied both intraday or on a day-to-day basis. There are often dynamics at play between indicators and they must be interpreted selectively and in combination.

In this paper, FA is not used, while TA is applied by using technical indicators as ANN input. Primary reasons for using technical indicators as input are the prevalence of this approach in literature (and encouraging results) and the belief that technical analysis possesses a larger predictive power than past. An example of a typical technical indicator – a price oscillator – is illustrated in Figure 1 below. A price oscillator is based on 2 moving averages (MA) of stock price; one with a longer period than the other. The shorter-period MA will "oscillate" around the longer-period MA. A typical interpretation is to consider it as a buy signal when the shorter-period MA rises above the longer-period MA. Conversely, when the shorter-period MA drops below the longer-period MA, a sell signal is generated.

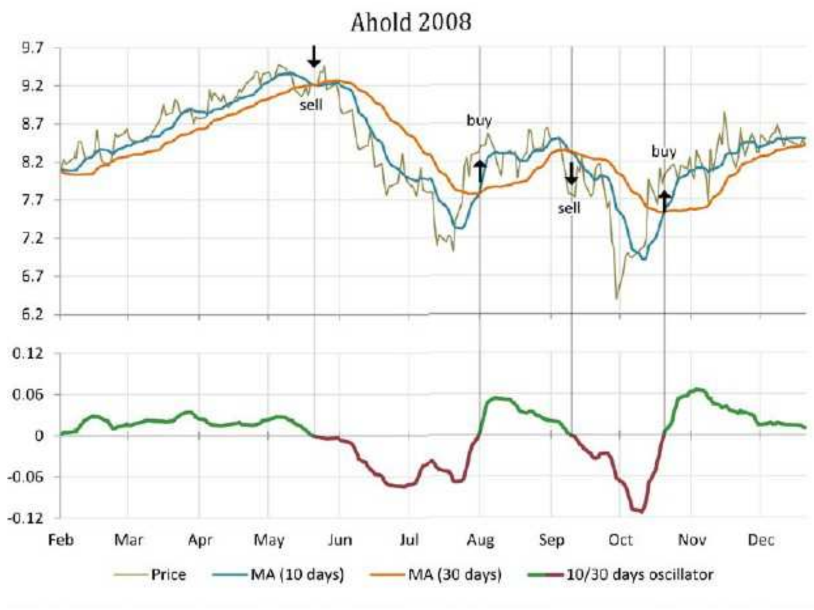


Figure 1: Price oscillator with 10-day and 30-day moving averages for Ahold in 2008.

Our TA/ANN intraday stock forecasting model employs a standard feed-forward ANN which is trained using the established back propagation training algorithm. Input data consists of technical indicators (see Table 1) based on raw observed stock data. During the training procedure, data is used to train the ANN to infer relationships between the input data and the desired outcome. This outcome is buying value, which represents a forecast of the stock's future movement. For the training procedure, the target value is pre-calculated for each input sample using future data samples. During forecasting, the ANN is expected to provide a target value at a specific point in time based on the current technical indicator input values, effectively forecasting future stock price movement.

An overview of the model is displayed in Figure 2 in the form of a data flow diagram. First, the application provides a data profile, which contains all settings regarding data, such as technical indicator configurations and size of the forecasting horizon. Based on these settings, a data set is constructed (processes 1 and 2), which is ultimately used to train and validate the ANN (process 3).

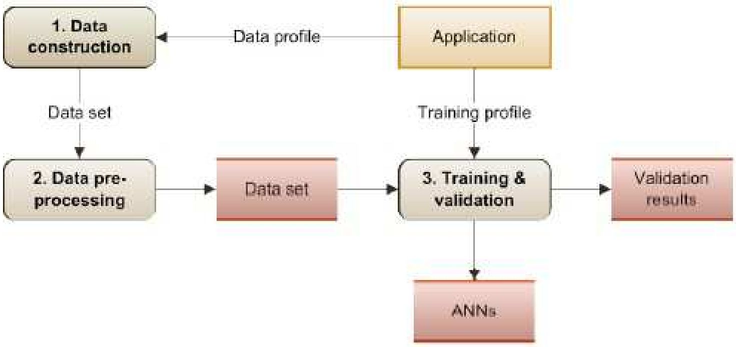


Figure 2: Model data flow diagram.

IMPLEMENTATION

The implementation has been realized on the Java SE 1.6 platform, mostly using the Eclipse IDE. Previously captured stock data was available on a MySQL database server. The MySQL Connector/J driver was used for Java-MySQL connectivity. The Standard Widget Toolkit (SWT; part of the Eclipse platform) was used in GUI development.

The entire system has been developed from scratch, with no use of libraries beyond SWT and the MySQL driver. During development, performance of the ANN component was verified against NSTOOLS to ensure a proper implementation.

The basis of the developed system lies in the abstract TimeSequence class and its relations (see the class diagram in Figure 3 below). It maintains a list of times, managed by the LongValues class, which is a self-expanding array of long values. All time values in the implementation are measured using long values, indicating the number of milliseconds that has passed since January 1, 1970. Subclasses DoubleTimeSequence and LongTimeSequence add a list of double and long values, respectively, to the times list of TimeSequence. The base class provides basic value management capabilities for code that is unaware of the specific subclass it is handling. The implementation at hand was chosen over a generics-based implementation or use of java.util.Vector<T>, because of the demands large data sets place on the system. DoubleValues and LongValues maintain arrays of their corresponding primitive types, internally, while a generics-based solution would rely on instances of Double and Long instead.

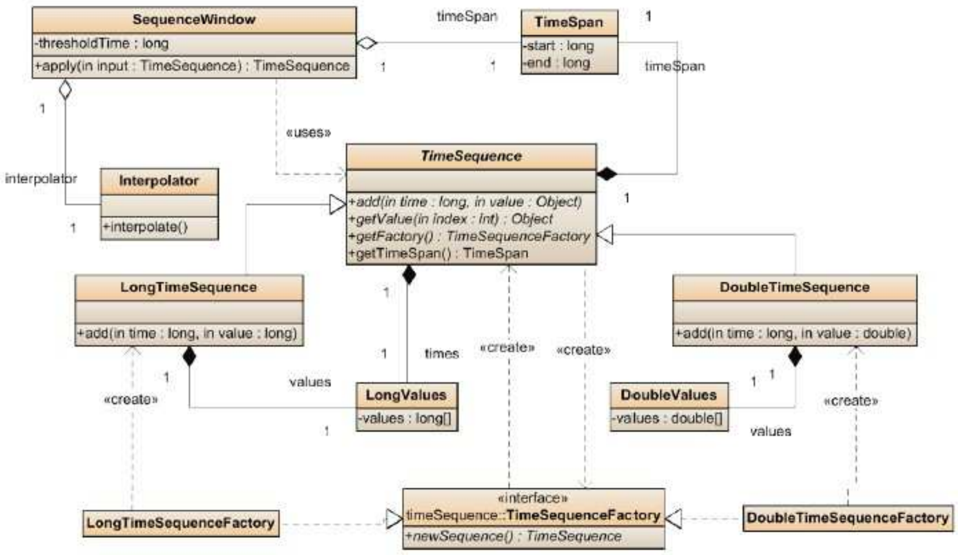


Figure 3: Class diagram of TimeSequence and related classes.

DATASETS

Several datasets are constructed for 3 different stocks and different configurations of technical indicators and buying value. The datasets are based on raw stock price and volume data, which was captured from a live data feed in the period February 2008 through December 2008. The selected stocks for forecasting are ING Groep N.V., Fortis N.V. and Koninklijke Ahold N.V. (ING, Fortis and Ahold for short). ING is a Dutch bank and insurance conglomerate, Fortis was a banking and insurance company as well (it is now defunct) and Ahold is an international group of supermarkets. The following motivates their selection. In 2008, ING and Fortis were in the top 5 of most traded stocks of the AEX index, which is the Dutch index, comprised of the 25 most traded stocks, thus providing large amounts of data to work with. All 3 stocks were relatively volatile in the considered time period, which results in a range of different situations to forecast. ING and Fortis show similar behaviour in price movement, which is important for the following reason: Technical indicator configuration is stock-specific; their parameters must be (radically) adjusted for individual stocks. The similarity in behaviour of ING and Fortis allows forecasting of both stocks using equal or similar parameters.

TRAINING ANN

To discover suitable combinations of ANN structures (topologies) and data profile for the three stocks for the ING, Fortis and Ahold a variety of ANN structures are trained and validated. The datasets are based on raw stock price and volume data, which was captured from a live data feed in the period February 2008 through December 2008. The selected stocks were at that time in the top 5 of most traded stocks. Each ANN structure was trained and cross-validated. ANN structures vary from 1 hidden layer with 12-32 neurons to several with 2 hidden layers. After preliminary experiments, a learning rate of 0.18 and momentum of 0.85 were adopted. The results of the experiment show a great variety with respect to the Error/performance measure SIGN, MAE and RMSE. The best performing combinations of ANN structures were used for additional experiments. The first experiment was to accurately train the selected models using early stopping. Also forecasting performance was enhanced and analysed further by applying a threshold. By using a threshold, cases for which no meaningful forecast can be provided are excluded, in turn increasing forecasting accuracy among the remaining samples. It proves that early stopping and setting a threshold showed somewhat equal responses. Finally the best performing ANN were selected. In Figure 4 we show one of the examples.

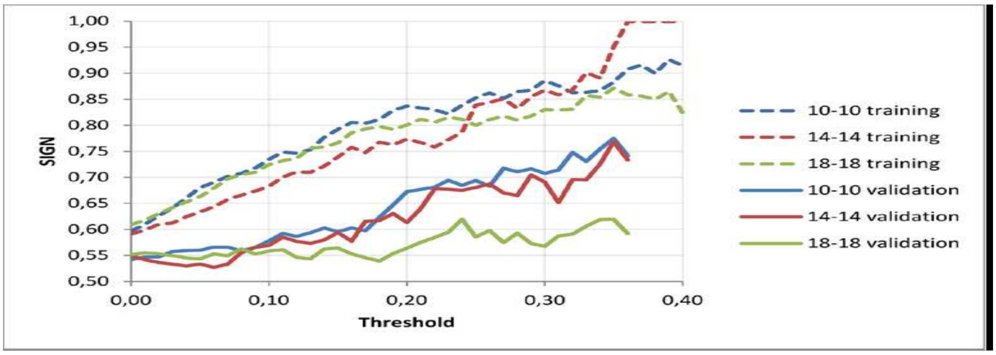


Figure 4: SIGN performance for ING forecasting in relation to the applied threshold.

TECHNICAL INDICATORS

Table 1: Technical indicator formulas

Technical indicator	Formula
Accumulation / distribution oscillator (A/D oscillator)	$f(t) = (H_t - C_{t-p})/(H_t - L_t)$
Commodity channel index (CCI)	$f(t) = (M_t - SM_t)/(0.015 \cdot D_t)$ $M_t = \frac{1}{3}(H_t + L_t + C_t)$ $SM_t = \frac{1}{n}\sum_{i=0}^{n-1} M_{t-i \cdot p}$ $D_t = \frac{1}{n}\sum_{i=0}^{n-1} \lvert M_{t-i \cdot p} - SM_t \rvert$
Disparity	$f(t) = C_t/MA_p$
Momentum	$f(t) = C_t - C_{t-p}$
Price oscillator	$f(t) = (MA_{short} - MA_{long})/MA_{short}$
Rate of change (ROC)	$f(t) = C_t/C_{t-p}$
Relative strength index (RSI)	$f(t) = 1 - 1/\left(1 + \left(\frac{1}{n}\sum_{i=0}^{n-1} Up_{t-i \cdot p}\right)/\left(\frac{1}{n}\sum_{i=0}^{n-1} Down_{t-i \cdot p}\right)\right)$ $Up_t = \begin{cases} C_t - C_{t-p} & \text{if } C_t > C_{t-p} \\ 0 & \text{otherwise} \end{cases}$ $Down_t = \begin{cases} C_{t-p} - C_t & \text{if } C_t < C_{t-p} \\ 0 & \text{otherwise} \end{cases}$
Stochastic fast %K	$f(t) = (C_t - L_t)/(H_t - L_t)$
Stochastic fast %D	$f(t) = MA_q(\%K)$ This is a moving average with period q of stochastic fast %K.
Stochastic slow %D	$f(t) = MA_r(\%D)$ This is a moving average with period r of stochastic fast %D.
William's %R	$f(t) = (H_t - C_t)/(H_t - L_t)$

TRADING SIMULATION

For each considered stock, a trading simulation is performed for the combinations of data profiles and ANN structures. The objective is to obtain a performance measure of the stock forecasting model in terms of obtained return on investment. Performance is compared to the buy & hold strategy and a naive forecaster / strategy. The simulations are performed on data, which was not involved in the training and cross-validation process. During a time period of two months the selected stock from ING, Ahold and Fortis is bought and sold based on forecasts from the TA/ANN model. At the end of each trading simulation, the obtained compound return is used as the resulting performance measure. Trading simulation performance of the TA/ANN forecasting model is compared with the simple buy & hold and NAIVE- strategies. Buy & hold buys the stock at the beginning of the period and sells it at the end. This serves as a measure of the stock's performance with no intermediate trading. When plotted as a graph, the return of a buy & hold strategy resembles the stock's price movement. This is because it is effectively the normalized price, where the initial price is equal to a buy & hold return of 1 (100%). The NAIVE- strategy generates buy and sell signals based on the last stock price movement: If the last stock price was an increase, a sell signal is generated. If the last stock price was a decrease, a buy signal is generated.

RESULTS

Table 2 below provides trading simulation results for each stock, listed per TA/ANN structure and the buy & hold and NAIVE- strategies. The total return is the compound return and provided as a fraction of the original investment. Thus, a total return of 1.15 would indicate a profit of 15%, while a total return of 0.87 would correspond to a 13% loss. Furthermore, the number of trades is listed under # trades. This corresponds to the number of stock positions held during the period. Thus, buying a stock and selling a stock are counted as a single trade. The return/trade is the return per trade, calculated from the total return and number of trades by taking the n-th root of the total return, where n is the number of trades. This may be interpreted as the quality of the performed trades. Finally, the average duration of stock positions is listed. This is the amount of time a stock was in possession (the time between buying and selling a stock). For each stock and forecaster / strategy, the obtained returns are plotted as a function of time in the considered period from early October through December 2008.Figure 5, 6 and 7, for ING, Fortis and Ahold, in that order. In the cases of ING and Ahold, the buy & hold graph falls largely outside the plot area: Properly plotting the returns for the TA/ANN structures was given priority.

Table 2: Trading simulation results per stock and TA/ANN forecaster or strategy.

Stock	Forecaster	Total return	# trades	Return/trade	Avg. pos. duration
ING	18-18	**1.21**	119	1.00160	1:11:06 hrs.
	14-14	1.11	120	1.00089	1:13:06 hrs.
	10-10	1.09	99	1.00090	1:20:20 hrs.
	Buy & Hold	0.51	-	-	-
	NAIVE-	1.07	1.236	1.00006	0:06:34 hrs.
Fortis	18-18	1.52	189	1.00222	0:53:04 hrs.
	14-14	**1.53**	196	1.00217	0:37:36 hrs.
	10-10	1.32	166	1.00168	0:59:34 hrs.
	Buy & Hold	0.77	-	-	-
	NAIVE-	1.31	816	1.00033	0:09:11 hrs.
Ahold	10-10	**1.25**	177	1.00125	0:48:37 hrs.
	24	1.21	181	1.00106	0:45:52 hrs.
	12	1.16	154	1.00098	0:58:08 hrs.
	Buy & Hold	1.10	-	-	-
	NAIVE-	1.14	855	1.00015	0:07:33 hrs.

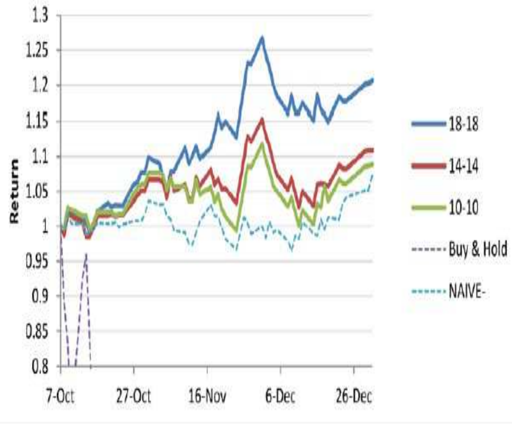
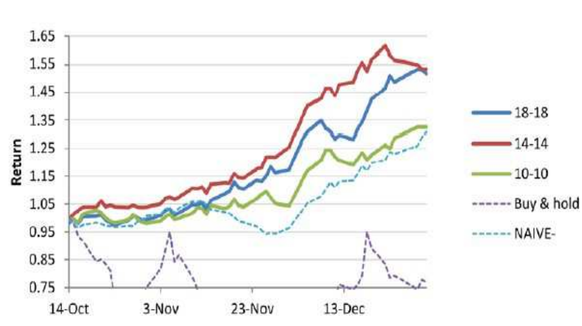


Figure 5: ING trading simulation returns for selected TA/ANN forecasters and simple strategies.



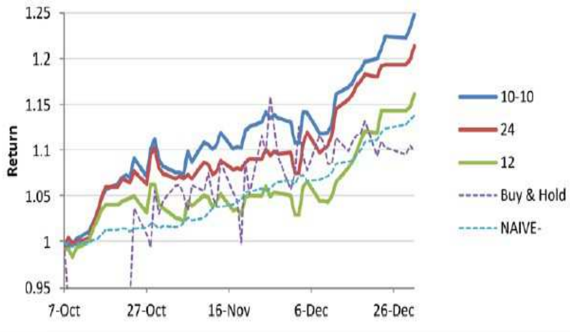


Figure 6, 7: Fortis/Ahold trading simulation returns for selected TA/ANN forecasters and simple strategies.

The simulation experiment performed trading simulations for each stock in order to evaluate forecasting performance in terms of obtained compound trading returns. Returns were compared to a buy & hold and naive strategy. The simulated trading period spanned over 2 months. After applying optimal thresholds, positive and negative buying values were interpreted as buying and selling signals, respectively. The best performing TA/ANN models obtained compound returns of 121% (21% profit) for ING, 153% for Fortis and 125% for Ahold. The models outperformed buy & hold and the naive strategy at virtually all times. For each stock, a single TA/ANN structure outperformed the others. For ING and Ahold, the best performers were the most complex structures (18-18, 10-10).

CONCLUSIONS AND FUTURE WORK

It proves that the used technical indicators and their parameters do not have a strong relationship with the target value. Of course ANN cannot infer a relationship that is not present. Alternative explanations for the weak relationship are that the data set is not large enough to capture all aspects of the complex, underlying system and a large part of data consists of conflicting samples, preventing the ANN from modelling the relationships present in the remaining data. But our simulation experiment demonstrated promising returns for the TA/ANN models in trading simulations. This supports the findings that the SIGN performance measure is more indicative of a stock forecaster's trading performance than the traditional performance measures. Next future we will research time series to forecast intraday stocks.

REFERENCES

[1] Zhang, G., Eddy Patuwo, M. Y. Hu (1998): "Forecasting with artificial neural networks: The state of the art", International Journal of Forecasting, Elsevier, 1998.

[2] Adya, M., F. Collopy (1998), "How Effective are Neural Networks at Forecasting and Prediction? Journal of Forecasting", Elsevier, 1998.

[3] De Gooijer, J.G., R.J. Hyndman, "25 years of time series forecasting. International Journal of Forecasting", Elsevier, 2006.

[4] Lee, R.S.T., iJADE "Stock Advisor: An Intelligent Agent Based Stock Prediction System", Man and Cybernetics, Part A, IEEE Transactions on, 2004.

[5] Tsang,P.M., P. Kwok, Design and implementation of NN5 for Hong Kong stock price forecasting, Engineering Applications of Artificial Intelligence, Elsevier, 2007.

[6] A Gijsberts, A., G Metta, L Rothkrantz , Evolutionary optimization of least-squares support vector machines, Data mining, Springer, 2010

[7] Dikkers,H., L Rothkrantz Support vector machines in ordinal classification: An application to corporate credit scoring Neural Network World, 2005

[8] Gerssen,S.G., L. J. M. Rothkrantz, Bayesian Network Learning for rare events, International Conference on Computer Systems and Technologies, 2006

ABOUT THE AUTHORS

Victor Louwerse, Ir,Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

Prof. Leon Rothkrantz, PhD, Multimodal Communication, Intelligent Interaction, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, E-mail: L.J.M.Rothkrantz@TUDelft.nl.