The Egyptian Stock Market Return Prediction: A Genetic Programming Approach

MOHAMMED E. EL-TELBANY

Computers and Systems Department
Electronics Research Institute, Dokki
EGYPT
Email: telbany@eri.sci.eg

Abstract:- Applications of learning algorithms in knowledge discovery are promising and relevant area of research. It is offering new possibilities and benefits in real-world applications, helping us understand better mechanisms of our own methods of knowledge acquisition. Genetic programming as learning algorithm posses certain advantages that make it suitable for forecasting and mining the financial data. Especially the stock time series have a large number of specific properties that together makes the prediction task unusual. This paper presents the results of using genetic programming to forecast the Egyptian Sock Market return. Experiments results demonstrate the capability of genetic programming to predict accurate results, comparable to traditional machine learning algorithms i.e., neural networks.

Key-words:- Genetic Programming, Neural Networks, Egyptian Stock market and Prediction.

1 Introduction

Recently, there has been upsurge of interest in the area of *data mining* in knowledge discovery. Traditionally, such attempts have been on prediction; by identifying the near future price movements. Financial data forecasting is one of the fundamental activities in data mining and the using of genetic programming (GP) for financial data forecasting is relatively under-explored area. It is a promising approach due to their effectiveness in searching very large spaces and the ability to perform global search for best forecasting model. Genetic programming can be applied to forecasting the Egyptian Stock Exchange (ESE), which is one of the oldest in the world and comprises of two exchanges, respectively the Alexandria and Cairo stock exchange. The behavior of the ESE stock return is evaluated using the CASE30 daily index, which measures the return on investment from the change in value of the stocks. The index tracks the performance of the most active 30 stocks and measures the return on investment from the change in value of the stocks. ESE stock return time series have a large number of specific properties that together makes the prediction task unusual. The process behaves very much like a random-walk process and regime shift in the sense that the underlying process is time varying and volatility in the time series change as the stock markets moves. These reasons cause greats problems for the more traditional and popular statistical algorithms for time series predictions such as ARCH and GARCH [8] [1] and the like which required a stationary time series with normality inputs and independence of the residuals. Genetic programming predictor as an alternative adaptive method to forecast the ESE stock return since it perform a *global* search in which genetic operators can modify many different combinations of variables - using the several different functions available in the function set. Hence, even if the original variables do not have much predictive power by themselves, the system can effectively create "*derived variables*" with greater predictive power, by applying the function set to the original variables [7] [2]. Finally, it offers wide range of possibilities of variations and modifications of algorithm that may lead to improve overall performance of the application. In this paper, genetic programming technique is used for forecasting the EST stock return and evaluates its predictability given the enormous amount of noise in financial markets. The GP technique is used for automatic discovery and the evaluation of forecasting model' performance and its predictability for the Egyptian stock market given the enormous amount of noise in financial markets represented is presented. Especially, the Egyptian stock market return data have a large number of specific properties that together makes the generalized forecasting model unusual. The process behaves very much like a random-walk process and regime shift in the sense that the underlying process is time varying. The rest of the paper is organized as follows. Section 2 presents related work. Section 3 presents an overview of genetic programming.

Section 4 examines the data used to assess the Egyptian stock market. Section 5 describe in detail the methodology used in this study. Section 6 presents the experimental results. Section 7 is a discussion of the results and concludes the paper.

2 Related Work

GP have been applied to financial time-series prediction by various authors since their expansion form the genetic algorithms. There are many works on application of using GP in finance and time series prediction. Koza [3][4] presents the first study using GP to the task of predicting next values of the logistic equation. The results of this short-term prediction were expectably good. Oakley [11] and Mulloy *et al.,* [9] study the using of GP in chaotic time series prediction. Mahfoud and Mani [6] presented a new genetic-algorithm based system and applied it to the task of predicting the future performance of individual stocks; Neely *et al.* [10] applied the GP to foreign exchange forecasting and reported some success. Li and Tsang [5] present a GP based system that takes a well-known technical rules and adapting them to stock prediction problems. Rafat *et al* [12] presenting the using of GP in technical trading rule induction for the Egyptian stock market.

3 Overview of Genetic Programming

Genetic programming (GP) [3][4], belong to a class of optimization techniques broadly called evolutionary algorithms. Genetic programming is modification of genetic algorithms with one major difference. The population consists of individuals represented by specific data structure – trees. Inner nodes of the trees can represent functions (e.g. arithmetic operators, conditional operators or problem specific functions) and leaf nodes would be terminals – external inputs, constants, and zero argument functions. Both the function set and the terminal set must contain symbols appropriate for the target problem. Each individual of the population is evaluated with respect to its ability to solving the target problem can grow in size and shape in a very dynamical way. This evaluation is performed by a fitness function, which is problem dependent function and the population evolved through the actions of genetic operators such as reproduction, mutation and crossover. Mutation means the evolution of a completely new structure determined at random at selected node (i.e., subtree in the chosen node is deleted and new one is randomly created, by "grown" in the chosen mutation point). Crossover involves the branches from two parent structures being swapped as shown in Figure 1.

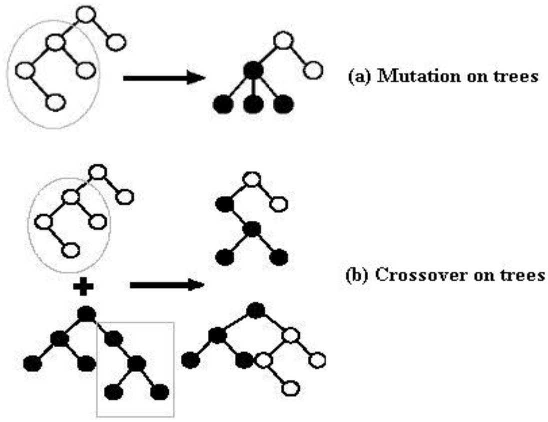

Fig. 1. : Genetic programming operators (a) mutation (b) crossover.

4 Egyptian Stock Market Domain

The Egyptian Stock Exchange (ESE) comprises of two exchanges, respectively the Alexandria stock exchange and Cairo. The two exchanges were integrated and govern by the same board of directors (i.e., Capital market authority) as an independent regulatory agency. The two exchanges share the same trade, clearing and settlement systems, so that market participates have access to stock listed in both exchanges. The behavior of the ESE stock return is evaluated using the many daily aggregated indices. The CASE30 daily index that is the most widely known and acknowledged performance indicator that measures the return on investment from the change in value of the stocks. The index tracks the daily performance of the most active 30 stocks from January 2001 to September 2003. The CASE30 index measures the return on investment from the change in value of the stocks. This gave us 663 data point; the data point presented the daily close price value $y(t)$, which the price of the last performed trade during the day for each stock is used to produce the prediction of the future values. However, this data is seldom to be used due to the prices $y(t)$ normally vary greatly and it is difficult to create an accurate model for a longer period of time. So that, the *return* as defined in Equation (1) is used instead.

$$R(t) = \frac{y(t) - y(t-1)}{y(t-1)} \qquad \textbf{(1)}$$

where $R(t)$ is the one-step return and represent the relative increase in the price since the previous point in the time series. We somewhat arbitrarily took the first 940 data points for training, the next 470 points for validation and test. Each data point comprises the stock return of the previous five days and the prediction of the future stock return.

5 Methodology

The stock time series prediction problem can be formulated as follows: Given a set of N examples, $\{(R_i, t_i),\ i = 1, \ldots, N\}$ where $f(R_i) = t_i, \forall i$, return a function g that approximates f in the sense that the norm of the error vector $E = (e_1, .., e_N)$ is minimized. Where each e_i is defined as $e_i = e(g(R_i), t_i)$ and e is an arbitrary error function. The system used follows the standard GP framework, where the GP expresses the predicted stock value as function in the previous stock values. GP in general has a five components, i.e., the function set, terminal set, fitness function, control parameters and stop condition, must be determined to solve a problem. The function set consists of the arithmetic and trigonometric functions

$$\{+, -, *, \%, \sin, \cos, \log, \exp\}$$

and the terminal set consists of all input return of the last week and ephemeral random constant ranges from $[-1.0, 1.0]$

$$\{R(t-1), R(t-2), R(t-3), R(t-4), R(t-5), \Re\}$$

Fitness function is determined by comparing the values returned by the GP tree to each value over the 940 fitness cases (i.e., the 940 point training set). The test set is used to evaluate the performance after the GP has found its best tree. The raw fitness, *r*, is the sum over the fitness cases of the squared error between the actual stock market return and the value predicted by the GP as follows:

$$r = \sum_{t}^{N} \left(R(t+1) - g(\mathbf{R}(t))\right)^2 \qquad (2)$$

The parameters we use are the population size is 5000; the number of generations is 101 (including the first random population). The initial population is generated using the *ramped-half-and-half* method. The maximum depth of the new individual is 6. The maximum depth of individual after crossover is 10. Fitness proportionate selection is used. 90% of selected individuals undergo crossover, with 70% internal and 20% external crossover points being selected. The remaining 10% of selected individuals undergoes reproduction.

6 Experimental Results

This section demonstrated the results of the GP runs, which explore various aspects of specification search. For each GP run, the fittest S-expression is evaluated over the training and test set using two measures of forecast performance: mean square error and mean absolute error (MAE). The results obtained from the GP are compared with the neural networks model results. Where a three-layer feed forward net consists of one input layer, one hidden layer containing six processing elements, and a linear output layer to improve the extrapolation properties. The network is trained by back-propagation training algorithm with learning rate 0.05 and number of iterations 30000.

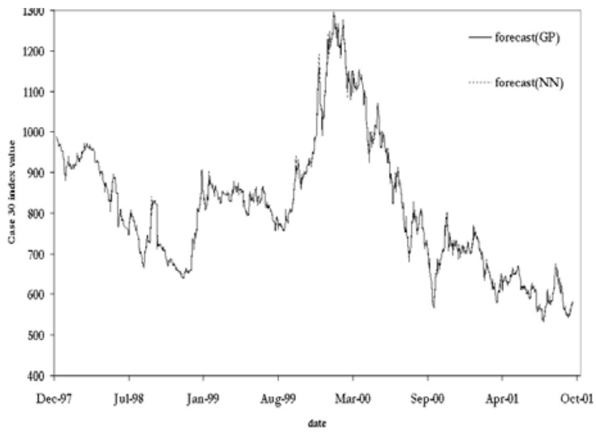

Fig. 2. Predicted Egyptian stock return for training data set by GP and NN models.

The predicted for training and test data set for GP and NN models are displayed in Figure 2 and Figure 3. The solid line corresponds to the predicted stock return using GP model and the dashed line to the predicted stock return using NN model.

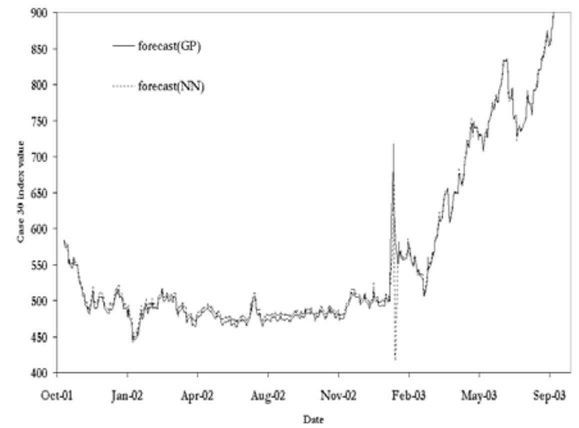

Fig. 3. Predicted Egyptian stock return for testing data set by GP and NN models

The forecasted and actually measured values where compared to verify the generated model by GP and NN. The results of a preliminary run on this Egyptian stock return data set are summarized in Table 1.

Table 1. The comparsion between forecasted and actual measerd stock return for GP and NN models.

	Training Set		Test Set	
	RMS	MAE	RMS	MAE
Genetic Programming	172.2	9.3	57.8	5.3
Neural Networks	189.6	9.7	175.6	7.1

From the results listed in Table1; we can conclude that the GP solution was able to mimic this very complex and dynamically changing output variable quite well. Moreover, the predictability of GP model performed better than the NN model since the RMS and MAE less than the NN model.

7 Conclusion and Discussions

In this work, we have tested the GP learning technique on the problem of predicting the daily Egyptian stock market return. GP offers many advantages in comparison with the sophisticated statistical techniques such as ARCH and GARCH including its ability to evolve arbitrary complex models. Moreover, the experiments demonstrated that the GP is performed better than the NN.

References:

[1] Azab B., (2002). The Performance of the Egyptian Stock Market. MSc. Thesis, Business School, University of Birmingham.

[2] Freitas A. (2002). *A survey of evolutionary algorithms for data mining and knowledge discovery*. In: Ghosh A. and Tsutsui S. (Eds.) Advances in Evolutionary Computation. Springer-Verlag.

[3] Koza J. (1992). *Genetic Programming: on the Programming of Computers by Means of Natural Selection*. Cambridge, MIT Press.

[4] Koza, J., (1994). *Genetic Programming II: Automatic Discovery of Reusable Programs*. MIT Press, 1994.

[5] Li J., and Tsang E. (1999). Improving Technical Analysis Predictions: An application of Genetic Programming. In Proceedings of Florida Artificial Intelligence Research Symposium.

[6] Mahfoud S., and Mani (1996). Financial Forecasting Using Genetic Algorithms. Journal of Applied Intelligence, **10**(6), pp. 543-565.

[7] Masand B., and Piatetsky-Shapiro G., (1996). Discovering Time Oriented Abstractions in Historical data to Optimize Decision Tree Classification. In Angeline P., and Kinnear K., (eds.) Advances in Genetic Programming II, MIT.

[8] Mecagni M., and Shawkey M., (1999). Efficiency and Risk-Return Analysis for the Egyptian Stock exchange. Working Paper 37, the Egyptian Center for Economic Studies.

[9] Mulloy B., Riolo R., and Savit R., (1996). Dynamics of genetic Programming and Chaotic Time Series Prediction. In Proceedings of the Genetic Programming First Annual Conference.

[10] Neely C., Weller P. and Ditmar R. (1997). Is Technical analysis in the forign exchange market Profitable? A Genetic Programming Approach. In Proceedings, forecasting financial market: Advances for exchange rates, Interest rates and Asset management, London.

[11] Oakley H. (1994). Two Scientific Applications of genetic Programming: Stack Filters and Non-Linear Equation Fitting to Chaotic Data. In Kinnear K., (ed.) Advances in Genetic Programming I, MIT.

[12] Rafat S., El-Telbany M., Hefny H., and Bahnaswi A. (2003). Discovering The Classification Rules For Egyptian Stock Market Using Genetic Programming. In Proceedings of the 46th IEEE-2003 Midwest Symposium on Circuits and Systems, MWSCAS, 2003.