A solution for forecasting pet chips prices for both short-term and long-term price forcasting, using genetic programming

Mojtaba Sedigh Fazli, Jean-Fabrice Lebraty

A Solution for Forecasting PET Chips Prices for both Short-Term and Long-Term Price Forecasting, Using Genetic Programming

Mojtaba Sedigh Fazli[1], Jean-Fabrice LEBRATY[2]

ABSTRACT

Nowadays, forecasting on what will happen in economic environments plays a crucial role for managers to invest correctly on appropriate items. We showed that in PET market how a neuro-fuzzy hybrid model can assist the managers in decision-making [13]. In this research, the target is to forecast the same item through another intelligent tool which obeys the evolutionary processing mechanisms. Again, the item for prediction here is PET (Poly Ethylene Terephthalate) which is the raw material for textile industries and it is highly sensitive against oil price fluctuations and also some other factors such as the demand and supply ratio. The main idea is coming through AHIS model which was presented by M.S. Fazli and J.F. Lebraty in 2013 [13]. In this communication, the hybrid module is substituted with genetic programming. Finally, the simulation has been conducted and compared to three different models answers which were presented before. The results show that Genetic programming results (acting like hybrid model) which support both Fuzzy Systems and Neural Networks satisfy the research question considerably.

KEYWORDS:

Efficient Market Hypothesis, Financial Forecasting, Chemicals, Artificial Intelligence, Genetic Programming, Decision Support System, Hybrid Neuro Fuzzy Model.

1. INTRODUCTION

Innovation of Artificial Intelligence opened new horizons to Financial Forecasting issues. Unfortunately, there are a lot of financial managers who do not believe in forecasting but the method of AI tools which follow and predict the time series trends is still a hot issue in management and mathematics. We suppose that we can capitalize on the previous work in order to provide current decision maker in a specific field with an adapted decision support system. In this paper we want to answer the following research question "*How to forecast PET chips prices in short-time and long-time?*"

1.1 Why the AI methods are appropriate for this issue?

To handle this project, there are 2 major categories: one uses traditional methods, in this category there are 2 major methods named Fundamental Analysis and Technical Analysis. The second solution is to use novel tools such as AI tools. Due to the nature of price trends in stock markets which follow a chaotic process [1], the research seems to be compatible drawing on AI tools. A chaotic system includes two different parts: one is stochastic and another part is deterministic. When the market trend is not too noisy, the deterministic part will be more than 50%, in this case for the remaining part, obviously there are a lot of parameters which affect the price direction and fluctuations. Because of the variety of factors which control and affect the curve, it's considered that this part is stochastic and random. Here, our aim is to find a formula which will be able to determine the next day's prices.

The 70s decade was a start point for mathematicians in terms of applying the new mathematics, time series and even some advanced tools, such as Artificial Intelligence, to verify the forecasting ability of stock and other market prices. Today, the prices of chemicals which are used as raw materials in lots of industries usually are determined in stock exchange markets, or they directly depend on some other prices, which are determined in stocks such as oil price, exchange rate etc. Researchers have done a lot of tests and experiments on price information and stock exchange index in some countries such as USA, UK, Canada, Germany, Japan Turkey, India and etc.[5-8], to find existence or non-existence of defined structure in stock price information. At that time, the most important thing for researchers was to reject the Random Walk Hypothesis [2]. Stock markets are affected and surrounded by lots of extremely interrelated parameters such as economic, social, political and even psychological indicators [3]. These mentioned indicators interact with each other in a sophisticated manner; therefore it is normally very difficult and even some times impossible to forecast the fluctuations of price trends in stock markets.

There are lots of forecasting tools which are applied to this field in both traditional and modern techniques [4-7]. With the development of artificial intelligence, researchers and investors hope that the market complexities can be untied. Previously in 90s, there was

[1] *Doctorate Student , University of Montesquieu Bordeaux 4 , Mojtabafazli@yahoo.com*
[2] *Full Professor , Lyon3 University – IAE Business school – Lab. Magellan EA3713 , jean-fabrice.lebraty@univ-lyon3.fr*

a research conducted by Johnson and his colleagues [8] who identified a lot of potential uses for neural networks in financial institutions, corporate finance and investments. Over the last 20 years, the applications of AI tools in financial solutions have been increased dramatically.

2. Literature Review

In this section a review about the components of our research question will be presented. Firstly, the notion of price behavior on a chaotic market will be explained and as a consequence, a review of the main AI models and proposed tool which are currently possible to apply for this problem are discussed.

2.1 Efficient Market Assumption and Chaos Theory

Price behavior (especially stock price) is a challenging issue which researchers have always faced [9]. The main challenge is whether market price behaviors are predictable or not. Some researchers believe that prices do not follow a specific trend, rather act in a "random walk" and cannot be predicted at all [3]. They are mostly advocates of a hypothesis called "The Efficient Market Hypothesis (EMH)". It has been proposed in the Efficient Market Hypothesis that in an efficient market the opportunities for profit are discovered so quickly that they seem to be opportunities [8]. Therefore there are no advantages of exclusivity and thus negating its potential performance. There has been a sense of doubt and uncertainty about the validity of the EMH, and some researchers attempted to use neural networks and other intelligent tools to validate their claims [2].

Markets are, in general, chaotic and usually the market curve follows chaos attitudes. A modern approach to modeling nonlinear dynamic systems like the market price trend which is fully relevant is named "*Chaos Theory*". Chaos theory considers a process under the assumption that "part of the process is deterministic and another part of the process is stochastic" [1]. Chaos is a nonlinear process which appears to be random. Various theoretical tests have been developed to test if a system is chaotic (has chaos in its time series). The deterministic part can be characterized using regression fitting, while the random process can be characterized by statistical parameters of a distribution function [7].

2.2 Genetic Programming [10-12]

It is an evolutionary method used mostly for optimization problems and works based on genetic operations. Genetic programming is a little different from genetic algorithms, the task and the aim of GP is to be able to reproduce computer programs. Genetic Programming follows Darwin's theory of evolution and his famous phrase "survival of the fittest". There is a population of individuals who marry each other and reproduce the new generation. After passing the time and reproduction cycles, the produced items try to survive and just the best and the fittest one will survive [10, 11].

2.1.1 Generating a Random Population

According to Koza [11], there are three techniques to generate the random population called: Grow, *Full* and *Ramped-half and-half*. Here the third method is selected.

2.1.1.1 The Genetic Operations

The evolutionary process will start by applying fitness test to all the individuals in the initial random population. The new population is formed by applying three main methods: *reproduction, mutation* and *crossover*. After completing the new population (i.e. the same size as the old) the old population will be eliminated.

2.1.1.2 Mutation

This process is applied on one individual. It happens when the new generation faces a deadlock, and after applying all other operations, the fitness function does not achieve better result.

2.1.1.3 Reproduction

Reproduction is where a selected individual copies itself into a new population. It effectively works the same as surviving an individual into the next generation. According to Koza [12], normally 10% of the population is selected for reproduction.

2.1.1.4 Crossover

Crossover requires two individuals and generating two different individuals for the new population. Figure 3-11 describes the Cross over process. Koza uses crossover on 90% of the population. The crossover plays the most important role in this process, since it generates the source of new individuals. There are a few other evolutionary operations: *editing*, *mutation*, *permutation,encapsulation…*

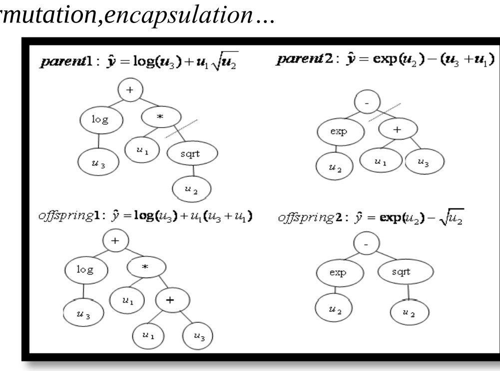

Figure 2-1: crossover genetic operation.

These methods would result in only a random search if it were not for the selection function.

2.1.2 Fitness-Proportionate Selection [9-11]

There is an algorithm which calculates the probability of selection, based on the nature of genetic programming and the definitions of this subject; it seems that the best individual of a population will be selected more frequently than the worst. The probability of selection is calculated with the following algorithm:

1. There is a raw fitness which will be restated in terms of *standardized fitness*. A lower value of standardized fitness denotes a better individual. By decreasing the individuals improvement through raw fitness, the standardize fitness will be equal to individual 'raw fitness. If the raw fitness decreases as an individual improves, standardized fitness for an individual is equal to the individual's raw fitness. And in case of increasing, an individual's standardized fitness is the maximum raw fitness minus the individual's raw fitness.

2. Standardized fitness is then reiterated as *adjusted fitness*, where a higher value indicates better fitness. The formula used for this is:

$$adj(i) = \frac{1}{1 + std(i)} \quad (1)$$

Where $adj(i)$ is the adjusted fitness and $std(i)$ is the standardized fitness for individual i. The application of using this adjustment is due to its benefits for separating individuals who have the near zero value of standardized fitness.

3. *Normalized fitness* is the form used by both selection methods. It is calculated from adjusted fitness in the following manner:

$$norm(i) = \frac{adj(i)}{\sum_{k=1}^{M} adj(k)} \quad (2)$$

Where $norm(i)$ is the normalized fitness for individual i, and M is the number of individuals in the population.

4. The probability of selection (sp) is:

$$sp(i) = \frac{norm(i)}{\sum_{k=1}^{M} norm(k)} \quad (3)$$

This can be implemented by [10]:
(a) Order the individuals in a population by their normalized fitness.
(b) Chose a random number, r, from zero to one.
(c) From the top of the list, loop through every individual keeping a total of their normalized fitness values. As soon as this total exceeds stop the loop and select the current individual.

2.3 AHIS Model [13]

The model used in this research previously was coined by us as AHIS. AHIS is an approach obtained from NORN which is presented by Ted Lee and colleagues on 2001[1]. But, finally the model is different from NORN due to some modifications which are applied for gaining more advantages and changes in this specific application. Moreover, some parts of that model are eliminated and we called it AHIS which stands for Adaptive Hybrid Intelligent System. This system makes the prediction stronger and more accurate in this specific application [13, 9].

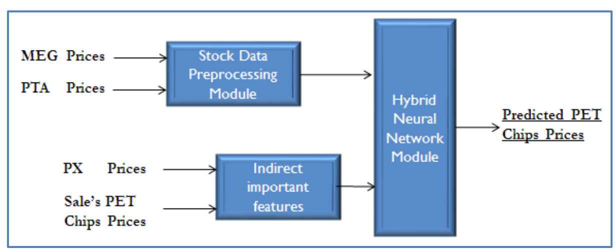

Fig 2 – 1: AHIS Model

As it was mentioned above there are 3 modules in AHIS model, in stock data preprocessing module, some preprocesses which are needed to be done on raw data take place. It is generally done for normalizing the data. Also, there is a module for applying indirect important features, which is normally effective in Neural Network methods, but here it's meaningless. In AHIS model PX price is applied as one of the most important features which indirectly affect PET chips prices. Another selected feature is sold PET chips prices, which is aimed to consider order and demand factors in the network. Since the cost price is produced by combining PTA and MEG in the first module through applying a specific formula. This formula is illustrated in Figure 2.2 [9]

Fig. 2 -2: Relationship between PX, PTA, MEG and PET chips

Finally, the last module is the most vital part of this model, where we previously tested pure neural network methods, such as a recurrent NN, MLP, TDNN, RNN, NARX. Besides that, some Neuro-Fuzzy models like ANFIS and LoLiMoT have been tested in our previous work. Here we changed our intelligent tool and used the Genetic programming approach instead of a Hybrid neural networks.

3. THE METHODOLOGY

3.1 Data

Input data are historical data of PTA, MEG. They are gathered through 2 reputed sources: one is ICIS which is well-known in statistics and the analysis of chemical market and another one in RECRON Company in Malaysia which is the biggest supplier of yarn in Asia. This issue is a big challenge in Asian yarn suppliers. The data set includes 347 price samples which are classified in 2 sub sets: one subset includes 247 samples which are used in training process and the remaining 100 are used in testing process for 1 step prediction. By increasing the prediction steps to 10 and 15 days, the training set size is increased and the test set is decreased. Random generation process follows Ramped half and half method [9].

3.2 Desired Prediction results criteria

Here there is a need to determine the acceptable error. In order to find a good idea in this issue, some in depth interviews have been done with experts in this field among East Asian chemical managers. Based on those interviews, the fitness factor and criteria could be explained as follows: If the error value which is the difference between real value and predicted value is lower than 80 USD/Ton the result is acceptable and fewer than 50 USD/Ton is desired. It means that such a difference is not very crucial on this market and will not have a big effect on the next item which will be produced from PET chips:

$$Err = |Fv - Rv| < 50 \qquad \textit{Desired} \ (4)$$

$$Err = |Fv - Rv| < 80 \qquad \textit{Acceptable} \ (5)$$

Err = Error Ratio,Fv = Forecasted Value,Rv = Real Value

3.3 Model

In G.P. based solution, regarding the suitable characteristics of Genetic Programming and the aim of this research, it seems that it’s possible to introduce a good and appropriate model for forecasting the mentioned items. However, simulating through Genetic Programming is a time-consuming process, it has been done in this research and the results were considerable .For running the simulation, the collected data are prepared in 15 days windowing .The data set is divided into two sets: first set (80% of total data) is used for training process and the second part (including 20% of total data) is used as test data. By considering the definition of the problem and our characteristic space of genetic programming, our research model will be changed to the following Model (See Fig. 3-1)

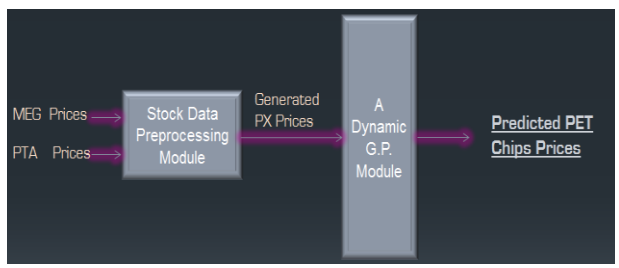

Figure 3-1: the research model is AHIS which is changed and modified to use for simulation of G.P.

As it is obvious in the above figure, the main module here is a dynamic G.P. module, running the G.P. in dynamic situation; it means that genetic operators are not fixed. In next step the trained data are used for modeling and the test one is used to check the validation accuracy, the used function set in simulation is as follows:

Function Set=(X^2,X^3,X^4,X^5,X^6,X^y,e^x,log(x), ln(x),$\sqrt{x}$, sin(x) , cos(x) , +,-,*, /)

For terminal set also following set is already used in this research:

Terminal Set = (rand , 1, 2, 3, 4, 5, 6, 7, 8, 9,10 ,11 ,12 ,13 ,14 ,15 ,16)

Here our logic obeys the fact that if we assume X_t for price of t^{th} day thus the next day price could be formulated as:

$X_{t+1} = X_t + \delta_{t+1}$(6)

Where δ could be a positive or a negative value. On the other hand, if we consider X_1 for the first day, other days prices predictions will be calculated as:

$X_2 = X_1 + \delta_1$ (7)
$X_3 = X_2 + \delta_2 = X_1 + (\delta_1 + \delta_2)$ (8)
$X_4 = X_3 + \delta_3 = X_1 + (\delta_1 + \delta_2 + \delta_3)$ (9)
$\vdots$
$X_n = X_{n-1} + \delta_{n-1} = X_1 + (\delta_1 + \delta_2 + \ldots + \delta_{n-1})$ (10)

In (10) we assume $(\delta_1 + \delta_2 + \ldots + \delta_{n-1}) = \delta$, therefore the i^{th} prices would be determined through the following function:

$F(X_i) = X1 + \delta$ (11)

Here, we try to find $F(X_i)$ through Genetic Programming and determine the δ function.

4. Results and Discussions:

After designing the model, simulations are conducted in Genetic Programming approach; also there were 3 different types of neural networks and neuro-fuzzy hybrid system which were presented by us before [9]. In the following simulations, around 247 patterns are considered in training sets and the remaining 100 samples are used for test set, so that the simulation is validated for the next 100 days, but in just 1 step prediction. All the results are gathered in 1 picture for doing a comparison.

Here, after simulation through G.P. with specific adjustment and setting, the following formula tree is generated (Fig. 4-1) and also the result for 1 step prediction is shown respectively in fig 4-2:

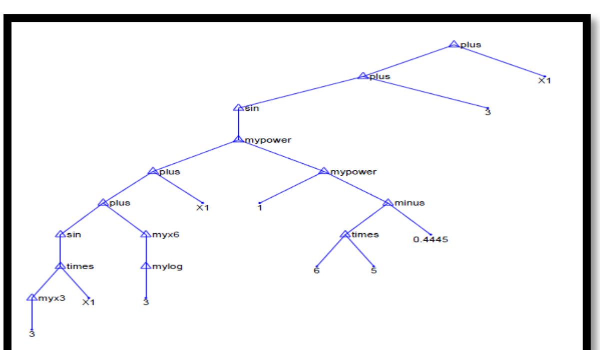

Figure 4-1: The Generated formula for finding the et days prices of PET with specific settings.

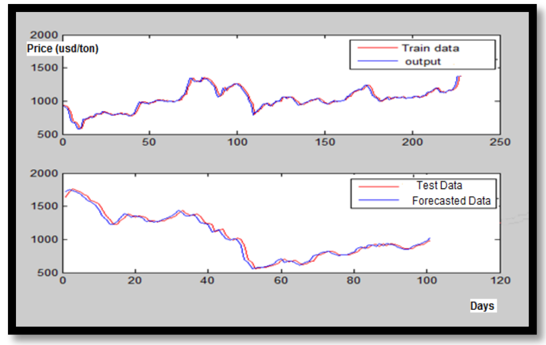

Figure 4-2: Forecasting using Genetic programming

The probability function for Mutation and Cross Over in simulation are not fixed and they have been adjusted dynamically and adaptively to obtain the best possible result (see fig. 4-3).

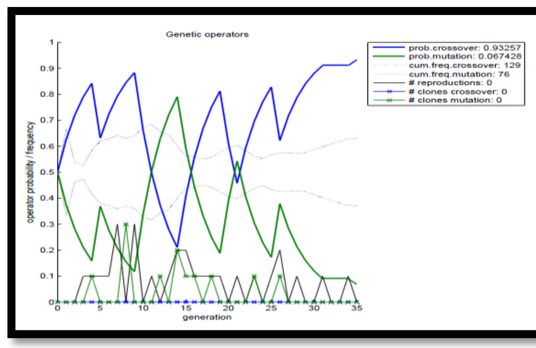

Figure 4-3 : operation probability and frequency is adjusted dynamically.

In the first step , the simulations were conducted for 1 step prediction and all the results were gathered and shown in figure 4-4 .It's obviously clear that the first two pure neural networks results cannot satisfy the research question, but the LoLiMoT and G.P. are competing to present the best possible answers.

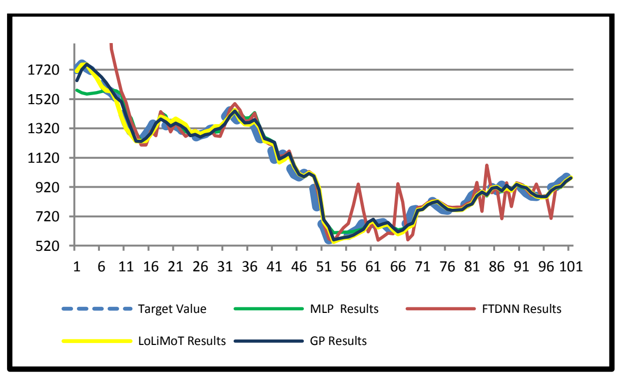

Figure 4 -4: All 4 simulations results for 100 days prediction with 1 step prediction in 1 figure

Generally, to check the error volume in such a problem, researchers use Normalized Mean square Error which is defined as follows:

$$\mathbf{NMSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i*=1}^{n} y_i^2} \quad (12)$$

Based on the above formula, the error rate of each model for 1 step prediction in LoLiMoT and G.P. are as follows:

MODEL NAME	MLP	FTDNN	LoLiMoT	G.P
NMSE RATIO	0.002831	0.07701	0.001411	0.001367

Table 4 -1: NMSE ratio for around 100 day's prediction with 1 step prediction for LoLiMoT and G.P.

Based on desirability which was defined previously, just the answers of these two models are desirable. In the next step, just these two models were tested for 5, 10 and 15 steps prediction. Figure 4-4 shows the results for 15 steps prediction for the last two models which already had better estimation .As it was mentioned in table 4-2, the error ratio for AHIS model including Neuro-fuzzy is absolutely better than other 2 models; However, G.P. results are even better in 1 step prediction (although the results are too close to each other). It means that for short-term predictions the G.P. wins the race. It seems that the number of patterns in this phase is not completely enough, but in this situation the results for LoLiMoT are better and ultimately the results are considerable. It means that for long-time prediction LoLiMoT answers are closer to our purpose. Table 4-3 demonstrates the results of 15 steps prediction for the next 15 days.

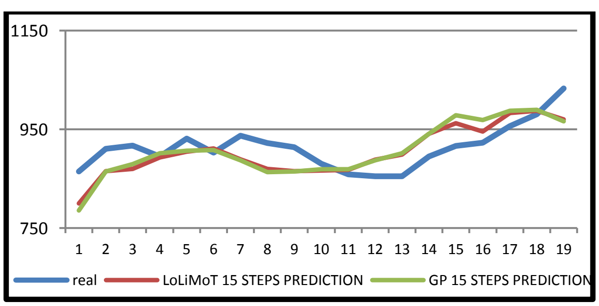

Figure 4 -4: selected simulation result for 19 days prediction with 15 step prediction for G.P and LoLiMoT

Test Day	Real Data	F.V. LoLiMoT	F.V. G.P.	Err. LoLiMoT	Err. G.P.
1	864.92	799.97	785.97	64.95	78.95
2	910.88	865.40	864.87	45.48	46.01
3	916.98	869.86	879.26	47.12	37.72
4	894.93	892.91	901.23	2.01	6.31
5	931.56	904.55	906.56	27.01	25.00
6	902.55	911.41	907.88	8.87	5.34
7	937.13	888.91	887.26	48.22	49.87
8	922.53	869.65	863.59	52.88	58.93
9	913.90	865.49	864.55	48.41	49.35
10	879.91	866.81	869.35	13.10	10.55
12	858.56	868.18	869.36	9.63	10.81
13	854.77	888.47	887.27	33.7	32.50
14	855.12	898.74	901.59	43.63	46.48
15	895.05	940.89	941.26	45.84	46.21

Table 4 -3: Forecasted value and Error volume for LoLiMoT with 15 step prediction

Finally, based on the formula which was mentioned in (12) the NMSE ratio would be as follows:

MODEL NAME	LoLiMoT	G.P.
NMSE RATIO	0.001859	0.002262

Table 4 -3: NMSE ratio for around 19 days prediction with 15 step prediction for both the LoLiMoT and G.P. answers

5. Conclusion:

In the introduction we asked "How to forecast PET chips prices for short time and long time?" This research follows our previous research which proposed a hybrid neuro-fuzzy system for predicting long-term forecasting in specific economic item. Previously, we showed that AHIS which includes LoLiMoT (a hybrid neuro-fuzzy model) provides a relevant answer to this question. Here, a G.P. based model has been tested and the answers for 1 step predictions improved the results of our previous research; however, still the results of AHIS including LoLiMoT are better in long-term forecasting. Here, the theoretical interest is to propose a new model that extends the Efficient Market Hypothesis. On the managerial Interest side, this model could be embedded in a Decision Support System (DSS). Our experience in that field indicates that such tools could be very useful for real decision-makers on PET market.

This communication has some Limitations. It seems that by increasing the number of testing samples and the range of training samples and events (especially in long-term prediction), the system would be more stable and the answers would be far more accurate. The last limitation is that, all other models which have the potential for better answers are not yet applied; Models such as using the Markov model and the combination of HMM with a neuro-fuzzy system. Therefore, for further researches it's strongly offered to researchers to find a model, combining the Markov Model with Neural Networks.

References:

[1] N. Masterakis, al., Recent Advances in Data Networks, Communications, and Computers, WSEAS Press, Bucharest, Romania, 2008.

[2] R.S.T. Lee, J.N.K. Liu, Norn Predictor – Stock Prediction using a Neural Oscillatory – based recurrent

network, International Journal of Computational Intelligence and Applications 1(2001) 431-459.
[3] Y.S. Abu-Mostafa, A. Atiya, Introduction to Financial Forecasting, Applied Intelligence, 6 (1996) 205-213.
[4] D. Zhang, J.J.P. Tsai, Machine Learning applications in Software Engineering, World Scientific Publishing, Singapore, 2005.
[5] J.W. Lee, P. Jonghun, O. Jangmin, L. Jongwoo, H. Euyseok, A Multiagent Approach to Q-Learning for Daily Stock Trading, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 37 (2007) 864-877.
[6] Y. Perwej, A. Perwej, Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm, Journal of Intelligent Learning Systems and Applications, 4 (2012) 108-119.
[7] M. Kaboudan, M. Conover, A Three-Step Combined Genetic Programming And Neural Networks Method Of Forecasting The S&P/Case-Shiller Home Price Index, International Journal of Computational Intelligence and Applications, 12 (2013) 1-16.
[8] B. Krose, P. Van der Smagt, An Introduction to Neural Networks, 8th ed., University of Amsterdam, Amsterdam, 1996.
[9] M.S. Fazli, Prévoir le prix des polyéthylène terephthalate (PET) en utilisant un
système intelligent, in, University of Montesquieu Bordeaux IV, Bordeaux, 2012.
[10] J.R. Koza, Genetic programming II: automatic discovery of reusable programs, MIT Press, 1994.
[11] J.R. Koza, Genetic programming: on the programming of computers by means of natural selection, MIT Press, 1992.
[12] J.R. Koza, D. Andre, F.H. Bennett, M.A. Keane, Genetic Programming III: Darwinian Invention and Problem Solving, Morgan Kaufmann Publishers Inc., 1999.
[13] M.S. Fazli, J.F. Lebraty,A Comparative Study on Forecasting Polyester Chips Prices, Using Different Hybrid Intelligent Systems, Accepted for publishing in IJCNN 2013 Conference Proceeding and IEEExplore, 2013.

APPENDIX I: LIST OF ABBREVIATIONS

AI Artificial Intelligence
ANN Artificial Neural Networks
DSS Decision Support System
EM Efficient Market Hypotheses
FA Fundamental Analysis
FDY Fully Drawn Yarn
FTDNN Focused Time Delay Neural
GP Genetic Programing
ICIS Integrated chemical information system
JSE Johannesburg Stock Exchange
LOLIMOT Locally Linear Model Tree
MEG Mono Ethylene Glycol
MLP Multi-Layer Perceptron
PET Poly Ethylene Terephtelate
POY Partially Oriented Yarn
PTA Purified Terephthalic Acid
PX Paraxylene
TA Technical Analysis