PREDICTION OF CURRENCY EXCHANGE RATE IN FOREX TRADING SYSTEM USING GENETIC ALGORITHM

Fauzi Yudhi Septiawan, Afiahayati
Dept. Of Computer Science and Electronics, Universitas Gadjah Mada

Chandra Kusuma Dewa
Department of Informatics, Universitas Islam Indonesia, Yogyakarta

In the foreign currency exchange (FOREX) system, there is always fluctuating price changes. Many people attempt to predict currency movements on the FOREX to be able to take appropriate decisions. Prediction method using a genetic algorithm with time-series data is one method that can be used in prediction problems. The genetic algorithm is an algorithm which inspired by the natural evolution. This study used genetic algorithm which begin by generate a population, perform roulette wheel selection method, followed by the whole process aritmethic crossover and random mutation. The results of the population will be selected by the method of steady-state updates. The calculation of the fitness evaluation was conducted using multiple linear regression with the number of variables that can be determined. Level prediction accuracy was measured by using the value of MAPD (Mean Absolute Percentage Deviation) and MSE (Mean Squared Error). This testing process use time-series data for each type of currency data which we want to predict for the span period of 1 year from January 2015 to December 2015. The tests were conducted on four types of major currency pairs in FOREX system that is USDJPY, USDCIF, GBPUSD and EURUSD. MAPD and MSE value which obtained on prediction from January to February 2016 for the USDJPY currency pair at 0,6930% and 1,1186533, USDCHF at 0,7425% and 0,0000734709051153, GBPUSD at 0,7190% and 0,0001600516436978 and for EURUSD at 0,4601% and 0.0001180997.

Keywords— forecasting, FOREX, genetic algorithm, time-series prediction

I. INTRODUCTION
Forex trading system can be called as currency transaction between countries. The value of currency in countries which always fluctuate against each other countries. This value is influenced by the strength of each other country [1]. Many researchers have studied...
various methods to make predictions on value exchange
rates in forex trading system [2][3][4][5][6].
Trader on forex system will always try to predict
the exchange rates to get profit from the difference
exchange rates and avoid the loss in transaction.
Therefore, many traders are looking for ways to
predict changes in currency values either by
fundamentals analysis that are directly related to the
overall economy conditions, industry conditions and
financial condition of the country concerned, or by
technical analysis by looking at the statistics generated
by market activity, such as past prices and volume of
the currency exchange value. If a trader can estimate
the strength of the particular currency exchange, of
course, he will be able to choose the option that
provides the greatest benefit and trying to avoid the
bad possibilities that occur when the value of the
currency exchange which he traded become weak.
Therefore, we need a system that can predict the value
of a currency with a pretty good degree of accuracy.
The methods which can be used in searching a
prediction solution is by using a genetic algorithm that
has a very promising computational models in
practical problems. Genetic algorithm method is an
algorithm of finding solutions which inspired by
natural selection process [7]. There is a group of
individual in a population that represents a set of
solutions that will be a subject to natural selection.
Individual that will survive is the one who has a good
solution. So the final solution is the best solution that
was obtained by natural selection.

II. RELATED WORKS
In a recent paper [2], the predictive value of the
currency exchange on the forex market is done by
using a genetic algorithm and SVR (Support Vector
Regression) + GHSOM (Growing Hierarchical self-
organizing map). GHSOM used for dividing each time
series into certain areas with almost the same
statistical distribution. While SVR used to predict each
region and also increase the strength of predictions. In
that study applied genetic algorithms to optimize the
four parameters of technical indicators so that later can
obtain the most favorable outcome. The results
showed that the genetic algorithm is far more superior
in predicting exchange rates in the forex system when
compared to SVR + GHSOM. Comparative value used
is a value gains (ROI) and maximum drawdown (MD).
In the other paper [3], used Feed Forward Neural
Network method which given training to form a net
with the best results. Data obtained from the
movements used in the form of a software MetaTrader
forex history data in hours. The test results prediction
are the open, low, high and close value which are
changing every hour. With a hybrid method of
artificial neural networks and genetic algorithm result
presentation RMSE level in Open value is 0.0071 with
95% accuracy rate, RMSE in Low value is 0.0328
with an accuracy of 50%. RM in High value is
0.0261 with a level of accuracy 59.17% and RMSE in
Close value is 0.0119 with an accuracy rate of 83.33%.
The difference with the research to be conducted by
writer is the author wanted to use genetic algorithms
with MLR as fitness evaluation function to the data in
the form of close-market price on the forex market
with 4 pairs of currency exchange rates of key
commodities.
In other paper [4], Hybrid method of artificial neural
networks and genetic algorithms are used to predict a
currency exchange rate with the US Dollar Peso
Philippine. However, they use fundamental analysis
that the input data used is the economic conditions that
occurred in the country concerned, namely in the form
of data demand from commercial banks which make
transactions through the system PDS (Philippine
Dealing System) and the supply is regulated by the BSP (Bangko Sentral ng Pilipinas). This transaction will take place every day and will be the currency exchange rate on that day. Neural networks are deployed using input variables such as consumer price index, the inflation rate, lending interest rate and the purchasing power of the peso. Although the data obtained is in the form of daily basis data, but the prediction results obtained are in the form of the average exchange rate in each month. The level of accuracy of prediction using neural networks is highly dependent on the number of sets of training data. So as to increase the level of accuracy, it is used hybrid genetic algorithm.

In his research related to the Forex market [5], used Hidden Markov Model method that retrieves data based on the pattern of the daily trend data for predicting the data in the next days. This model is divided into three stages, data analysis, training HMM using Baum Welch algorithm and predictions forward by using HMM. In the first stage in a linear regression is used to segment the general pattern uptrend and downtrend in Forex system. Next will be obtained extracts features that will be used for data input and as training for the movement trend. In the second stage will be used for training HMM with trend patterns that get results uptrend and downtrend. In the final stage, to predict the next trend patterns, then forward algorithm was used to evaluate the trained models.

In [6], they use genetic programming (GP) and gene-expression programming (GEP) to create an algebraic function that can be set up to represent the best prediction results. This research was conducted with the data changes in the market value of the stock in the short term or the long term. The data used in the form of time-series stock derived from five major companies renowned namely Yahoo, British Petroleum, Glaxo Smith Client (GSK), HSBC and RBS. The results obtained pattern accuracy in the short term is 5 days by 93.46% and for the long-term 56-day period with an accuracy of 92.10%.

In this study, Genetic Algorithm (GA) model is developed to predict exchange rate of JPY, CHF, GBP and EUR for unit USD using time-series data based on daily-time which collected from Jan 2015 to Des 2015 to predict the value exchange rate in Jan 2016 and Feb 2016. GA model can be used to get an adaptive slope and intercept value in multiple linear regression (MLR). This pattern function is used to make a next day prediction from a few days before.

III. FUNDAMENTALS

A. Genetic Algorithm

Genetic Algorithm is an evolutionary algorithms that has adaptive methods used to solve a search value in a real-world optimization problems and a variety of fields [2].

In GA, the decision space is referred as the environment. The potential solutions to the optimization problem are called chromosomes, which are solutions that represent a set of decision variables. The total number of solutions is called the population size and an iteration of the optimization process is called a generation. Reproduction, crossover and mutation are the essence of a standard genetic algorithm. [8]

1) Initialize Population

In the first step, the GA produces an initial population randomly, each consisting of some potential solutions. The decision variables are usually encoded as strings of binary digits or real numbers. There are variety type of encoding chromosome in GA. Encoding chromosome very depends on the problem itself. Figure 1 show some type of encoding chromosome representation in GA.
2) Fitness Function

The objective function values, called fitness function, are then calculated for each individual. For minimization problems, individuals with lower objective function values will have a higher probability of being selected for the next generation.

3) Crossover

New individuals are then created from parents of the current generation by crossover and mutation. Crossover includes cutting two pieces of chromosomes based on the desired number of points then combine half of each chromosome with other couples. It replaces old individuals in the population and are usually similar to their parents. Figure 2 show the simple 3 points crossover with binary representation.

4) Mutation

Mutations include the replacement value or changing value in random gen on population.

5) Elitism

The last part is elitism, using fitness evaluation to give a rank for each individuals, some of individuals are selected to become survivor for the next generation [2][8][9]. Figure 3 show the process on Genetic Algorithm.

B. Forecasting

Prediction or forecasting could be defined as the process of forecasting a variable or incidence in the future by the relevant variable data based on the past or based on specific data before. The data of the past systematically will be combined use of a particular method and processed to obtain forecasts on future circumstances.

Regression is a forecasting technique that measures the relationship of one variable to one or more other variables. For example, if we know that something has
caused product demand to behave in a certain way in the past, we might like to identify that relationship. If the same thing happens again in the future, then we can predict what demand will be [10].

In a simple linear regression, the model used one independent variable to make a relation with one dependent variable in form of a linear equation [11]. The equation for simple linear regression can be shown as in (1).

\[ Y = a + bx \]  

Where a is intercept, b is slope and x is independent variable in relation with one dependent variable y.

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y [11]. The equation for multiple linear regression can be shown as in (2).

\[ Y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \]  

Where a0 is intercept and a1 to an is slope or parameter from the independent variable x1 to xn.

C. Forex Trading System

The foreign exchange is the exchange of one currency for another, or the conversion of one currency into another currency. Foreign exchange also refers to the global market where currencies are traded virtually around-the-clock. The term foreign exchange is usually abbreviated as "forex" and occasionally as "FX" [12].

The differences in supply and demand within a certain period of time makes fluctuations in currency values compared to the other one. This differences in the value of money at a time is then utilized to take advantage. Finally currencies are traded in a market called the forex market.

The forex market includes trading between large banks in the world, central banks, major companies all over the world, governments and a wide range of other financial institutions.

IV. METHODOLOGY

In this study, i use GA to get the optimization of parameter pattern which used in Multiple Linear Regression (MLR) function as a slope value. Evaluation fitness can be got from this pattern as a Mean Squared Error (MSE) from difference error for each day data training. I will use MLR pattern to make a prediction on each day with system 1-day sliding windows. The best pattern result on each training will be compared to best pattern on data storage. If best pattern result training is more better than in data storage, then it will be stored in data storage. This best pattern for each type data will be used to predict the currency exchange by each type data. The results will be visualize by line graph or text report. This study use Java as programming language in NetBeans IDE platform. Figure 2 show the general process in this method using GA to optimize parameter in MLR.

Fig. 4. Scheme process of prediction system using GA.

A. Time-series Data

Time-series data obtained from FXDD-MetaTrader 4 which is a very popular software platform Forex Trading online and frequently used among traders [12][13]. Figure 3 show the data history center on MetaTrader4 system.
Fig. 3. Data history center based on daily data in MetaTrader4.

Data which taken from MetaTrader4 is data of exchange rate since the month of January 2015 to December 2015. Four type of currency exchange rate which obtained from MetaTrader4 is USD/JPY, USD/CHF, GBP/USD, and EUR/USD based on daily data in form of .csv file.

B. Genetic Algorithm Model

There are 6 steps in Algorithm Genetic method. These are initialize population, fitness evaluation, individual selection, crossover, mutation and elitism [8].

1) Initialize population

Initiate a random n-population with each of population represent a set of pattern where the variable of each gen on chromosome is defined as intercept and slope on MLR. It can be shown in (3).

\[ Y = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \ldots + \theta_n x_n \] (3)

where \( \theta_1 \) until \( \theta_n \) is each gen on chromosome and variable \( x_1 \) to \( x_n \) is value currency from data.

The chromosome representation using a real number representation which can be set random value at range 0 to 1. Gen length on chromosome can be set at least 3 value (MLR function has minimum 1 intercept and 2 slope or at least 3 variable which can representing as gen). Figure 4 show the example of chromosome representation.

![Chromosome representation with 5 gen value.](image)

2) Fitness evaluation

Evaluation fitness can be obtained from the value of 1/Mean Squared Error (MSE). The equation can be shown in (4).

\[ MSE = \frac{1}{n} \sum (y' - y)^2 \] (4)

where \( y' \) is prediction value, \( y \) is actual data from MetaTrader4 [12] and \( n \) is the total of data prediction. Prediction value can be obtained using MLR function (5).

\[ Y' = \theta_1 + \theta_2 (i-4) + \theta_3 x(i-3) + \theta_4 x(i-2) + \theta_5 x(i-1) \] (5)

where \( \theta_1 \) to \( \theta_5 \) is gen value in each of chromosome and \( x_1 \) is day data recent. Equation (5) using 4 day data recent to make a prediction data this day. If we want to predict the 5th day currency value then we use data from 1st to 4th day. In this study i use 3 to 7 gen length, which mean using 2 to 6 day data recent before.

3) Selection

Individual selection which become parent in crossover process can be obtained using Roulette-Whell method. First we calculate the fitness value of each individual \( (f_n) \). Second, we will calculate the value of the cumulative probabilities \( (P_n) \) for each individual

\[ P_n = f_n \sum f_n \] (6)

Third, generating random numbers which range 0 to 1 as much as \( n \) rounds. Fourth, selecting individuals where the random number selected as a parent based on the cumulative probability of each individual. Fifth, eliminating the parent index that have been selected so that the same individuals will not selected twice.

The number of individuals who are selected depending on the probability of crossover which be set at 50%, 60% and 70% from total of population.

4) Crossover
Crossover is the process of taking a part or allele sequence fragments from DNA genome to be combined to make new genome of a descendant or offspring. The crossover process performed using whole arithmetic crossover where the value of each gen in new child \(x'\) and \(y'\) can be shown as (7) and (8),
\[
x'_i = \alpha x_i + (1-\alpha) y_i, \quad 1 \leq i \leq n \quad (7)
\]
\[
y'_i = \alpha y_i + (1-\alpha) x_i, \quad 1 \leq i \leq n \quad (8)
\]
where \(x_i\) is gen on chromosome parent 1 and \(y_i\) is gen on chromosome parent 2, \(\alpha\) is a coefficient constant which can be inputted random between 0 to 1. In this study, the value of coefficient \(\alpha\) is 0.4. Figure 5 show the example of whole arithmetic crossover process in this study.

![Whole Arithmatic Crossover with coefisien \(\alpha\) at 0.4.](image)

5) Mutation
Number of mutation depending on the probability of mutation which be set at 0.5%, 1% and 2%. The point which gen mutated is selected random between 1st gen until the last gen on last chromosome. The value of gen which mutated is multiplied by random value 0.1 to 1.1.

6) Elitism
Elitism using steady-state update where the number of initial population will be keep at the same number so the population will not increase. The selection process will not change the overall individual with the new child individual. But only the individual with the best fitness values to be entered in the next iteration. This selection process will produce a number of individuals equal to the number of early individuals so it will not increase the number of individuals in the population.

The first step is to sort all existing individual based on the fitness value of each individual, and then take a number of \(n\) individuals who are equal to the number of initial individuals. Some individuals with the smallest fitness value from below will be deleted and just high fitness individuals will remain.

Figure 6 show the example of elitism process in this study.

![Elitism with 10 individuals on population.](image)

V. EXPERIMENTAL SETUP

1) Training Phase
Dataset used for this experiment is currency exchange rates of USD/JPY pair, USD/CHF pair, GBP/USD pair, and EUR/USD pair. Close value for a year data from Jan 2015 to Des 2015 are used in this training phase. Aim in this training is to get a MLR pattern using GA which has high fitness evaluation. This pattern will be used for prediction in data Jan 2016 to Feb 2016.

The training process conducted at several times with stopping condition at certain number of generations or iteration which has a best fitness evaluation. The chromosome which has best fitness value will be stored on best chromosome data storage for each type
of currency exchange. This best chromosome of each type data will be used on prediction phase.

Parameter setting for GA experiment is showed in Table I.

TABLE I. PARAMETER SETTINGS IN GA FOR EACH TYPE DATA

<table>
<thead>
<tr>
<th>No</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of genes</td>
<td>3 to 7</td>
</tr>
<tr>
<td>2</td>
<td>Population</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Crossover probability</td>
<td>50%</td>
</tr>
<tr>
<td>4</td>
<td>Mutation probability</td>
<td>0.5%, 1% and 2%</td>
</tr>
<tr>
<td>5</td>
<td>Stopping condition</td>
<td>50 to 500 generations (depending on each best combination results obtained)</td>
</tr>
</tbody>
</table>

2) Prediction Phase

In prediction phase, for each type currency data will use its best pattern chromosome which taken from best chromosome data storage. On USD/JPY, prediction will use best pattern from USD/JPY best chromosome in storage system. On USD/CHF, prediction will use best pattern from USD/CHF best chromosome in storage system. It also applied to GBP/USD data and EUR/USD data. So we have 4 type data best pattern on storage.

In this study, prediction will be conducted in Jan 2016 to Feb 2016. The accuracy of the prediction value can be measured using the value MAPD (Mean Absolute Percentage Deviation) which can be calculated as (9).

\[
MAPD = \frac{\sum|D_t - F_t|\Sigma D_t}{(9)}
\]

Predictions can be more accurate when the MAPD obtained close to 0%. The other accuracy can be measured by MSE (Mean Squared Error) which shown in (4). The smaller the MSE value obtained then predictions can also be said to be getting more accurate.

In other scenario we can set by ourself the n-following day time which prediction occur.

VI. EXPERIMENTAL RESULTS

In training phase, for each type data of currency exchange, we can get best chromosome that will be used in prediction phase. Table II - V show the partial results of training parameters, MSE value and fitness value which obtained at training phase.

TABLE II. PARTIAL RESULT OF TRAINING USING USD/JPY

<table>
<thead>
<tr>
<th>Gen</th>
<th>Pc</th>
<th>Iteration</th>
<th>MSE</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>100</td>
<td>0.52</td>
<td>120.32</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>100</td>
<td>0.43</td>
<td>120.32</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>100</td>
<td>0.67</td>
<td>120.32</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>100</td>
<td>0.78</td>
<td>120.32</td>
</tr>
<tr>
<td>5</td>
<td>50%</td>
<td>100</td>
<td>0.89</td>
<td>120.32</td>
</tr>
</tbody>
</table>

TABLE III. PARTIAL RESULT OF TRAINING USING USD/CHF

<table>
<thead>
<tr>
<th>Gen</th>
<th>Pc</th>
<th>Iteration</th>
<th>MSE</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>100</td>
<td>0.52</td>
<td>120.32</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>100</td>
<td>0.43</td>
<td>120.32</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>100</td>
<td>0.67</td>
<td>120.32</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>100</td>
<td>0.78</td>
<td>120.32</td>
</tr>
<tr>
<td>5</td>
<td>50%</td>
<td>100</td>
<td>0.89</td>
<td>120.32</td>
</tr>
</tbody>
</table>

TABLE IV. PARTIAL RESULT OF TRAINING USING GBP/USD

<table>
<thead>
<tr>
<th>Gen</th>
<th>Pc</th>
<th>Iteration</th>
<th>MSE</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>100</td>
<td>0.52</td>
<td>120.32</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>100</td>
<td>0.43</td>
<td>120.32</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>100</td>
<td>0.67</td>
<td>120.32</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>100</td>
<td>0.78</td>
<td>120.32</td>
</tr>
<tr>
<td>5</td>
<td>50%</td>
<td>100</td>
<td>0.89</td>
<td>120.32</td>
</tr>
</tbody>
</table>

TABLE V. PARTIAL RESULT OF TRAINING USING EUR/USD

<table>
<thead>
<tr>
<th>Gen</th>
<th>Pc</th>
<th>Iteration</th>
<th>MSE</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>100</td>
<td>0.52</td>
<td>120.32</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
<td>100</td>
<td>0.43</td>
<td>120.32</td>
</tr>
<tr>
<td>3</td>
<td>50%</td>
<td>100</td>
<td>0.67</td>
<td>120.32</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
<td>100</td>
<td>0.78</td>
<td>120.32</td>
</tr>
<tr>
<td>5</td>
<td>50%</td>
<td>100</td>
<td>0.89</td>
<td>120.32</td>
</tr>
</tbody>
</table>

Table II – V show the training results for each type data. From the table we can see that chromosome pattern with 3 gen has more higher average fitness value than the other amount of gen number.

Figure 7 show the results of training data in 1 year data from January 2015 to December 2015 for each type data which has best fitness value.
Chromosome in the best fitness will stored in best chromosome storage and it will be use to make prediction. The results prediction in Jan 2016 to Feb 2016 which obtained by using the best chromosome from the training data are showed by figure 8.

In figure 8 using USD/JPY data, the results prediction are obtained with the accuracy MAPD value at 0,6930% and MSE value at 1,1186533. Using USD/CHF data, MAPD value is obtained at 0,7425% and MSE value at 0,0000734709051153. In GBP/USD data, MAPD value at 0,7190% and MSE value at
0.0001600516436978. Using EUR/USD data, MAPD value is obtained at 0.4601% and MSE at 0.0000501171515238. This results show that for each type of data, MSE and MAPD value that has been obtained is on very low value. It means that the GA method in this study has high accuracy.

VII. CONCLUSION
The proposed forex prediction using GA model with MLR in time-series data offered a promising results for each type currency exchange rates.

In the training phase, the number of gen which have high chromosome is at 3 gen. Probability of crossover and probability of mutation make the GA to get the optimum point more faster or slower depending on each parameter in training phase. This can be showed at the number of generations which inputted in training phase to get the best chromosome.

In the prediction phase, the model show that the prediction has high accuracy in USD/JPY data, USD/CHF data, GBP/USD data and EUR/USD data. The results prediction showed that MAPD accuracy for all type data is very small, for USD/JPY data is 0.6930%, for USD/CHF data is 0.7426%, for GBP/USD is 0.7190% and for EUR/USD is 0.4601%.

It means that GA model using MLR as fitness evaluation is proved to be the promising method for prediction problem in forex trading system.

REFERENCES