Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Grey Relational with BP_PSO for Time Series Foreasting

Roselina Sallehudin
Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia
Skudai, 81300 Johor, Malaysia
roselina@utm.my

Siti Mariyam Hj.Shamuddin
Faculty of Computer Science and Information System,
Universiti Teknologi Malaysia
Skudai, 81300 Johor, Malaysia
mariyam@utm.my

***Abstract*—This paper proposes an effective hybridization of grey relational analysis (GRA) and Backpropagation Particle Swarm Optimization (BP_PSO) for time series forecasting. The hybridization employs the complementary strength of these two appealing techniques. Additionally the combination of GRA and BP as cooperative feature selection (CFS) has successfully assessed the importance of each input variable and automatically suggest the optimum input numbers for the forecasting task. Therefore it assists the forecaster to choose the optimum number of dominant input factor without a need to acquire expert domain knowledge. It also helps to reduce the interference of irrelevant inputs on the forecasting accuracy performance. To test the effectiveness of the proposed hybrid GRABP_PSO, the dataset of closing price from Kuala Lumpur Stock Exchange (KLSE) is used. The results show that the proposed model, GRBP_PSO out performed BP_PSO model and BP model in term of accuracy and convergence time.**

***Keywords*—Grey relational analysis, Backpropagation, Particle swam optimization, Time series forecasting, Coopeative feature selection, Forecasting accuracy.**

I. INTRODUCTION

Time series analysis and forecasting was used to forecast the developing trend or changes in the future according to a data set arranged by time. It has been applied widely in many different fields such as economics, sociology and science. Forecasting stock prices has been regarded as one of the most challenging applications of modern time series forecasting and very important for the success of many business and financial institutions. ARIMA model has been widely used for time series forecasting. ARIMA is a general univariate model and it is developed based on assumption that time series being forecasted are linear and stationary. Though, in reality most of the time series problems are complex and nonlinear. Therefore, using ARIMA model is not an appropriate solution for forecasting complex time series such as stock price prediction. Recently, nonlinear ANN model has been applied for stock price forecasting due to its promising performance in forecasting, classification and pattern recognition [1]. The ANN has the powerful capability to generalize the nonlinear relationships between inputs and desired outputs, without considering real problems domain expression. Both ARIMA and ANN need a large amount of historical data in order to yield accurate results. But, in a real world situation especially in stock market, the environment is full of uncertainties and changes occur rapidly, thus future situations must be usually forecasted using scant data available over a short span of time. Therefore forecasting in this situation requires a method that works efficiently with incomplete data. Although fuzzy forecasting methods are suitable for incomplete data situations, their performance is not always satisfactory [2]. Moreover, there are many affecting factors that influence the stock price simultaneously and usually these factors such as political events; general economics and trader's expectations are mixed together. As a result, the relationship between the stock prices and the affecting factors is ambiguous or the relational information is incomplete. Consequently, the selections of the affecting factors, which are relevant to modeling of input-output relationships, become a difficult task. Therefore in this study, grey relational analysis and BP neural network will be adopted as a cooperative feature selection to filter out the most effective factors when dealing with incomplete data situations.

Until now Backpropagation (BP) neural network is one of the widest application networks in time series forecasting. BP is suitable for time series forecasting because both of them are regarded as supervised learning. However, BP algorithm relies on the selection of initial weights and threshold that is normally chosen randomly. Thus it will easily get stuck in a local minima and converge slowly [3]. Subsequently, the network training became inconsistent and led to unreliable forecasting results. To overcome this drawback, PSO is employed in this study to determine the initial weights and threshold of BP. In order to overcome the incomplete time series data that contains irrelevant factors and to reduce the local minima problems, a grey relational PSO-based on BP neural network model (GRBP_PSO) is developed for forecasting time series data. To validate the performance of GRBP_PSO, daily Kuala Lumpur stock exchange closing prices are used. The experiment results show that the proposed model outperformed the BP model and BP-PSO model. The remainder of this paper is organized as follows. Section II presents the related literature review. Section III explains the implementation of the proposed model. Section IV discusses on experimental data and the obtained results from the study. Finally, Section V provides the summary and conclusions.

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

II. RELATED LITERATURE REVIEW

A. Grey Relational analysis(GRA)

GRA is a method of analysis, which has been proposed in the Grey system theory and founded by Professor Deng [4,5]. The purpose of GRA is to measure the relative influence of compared series on the reference series. In other words, the calculation of GRA reveals the relationship between two discrete series in a grey space.
There are three (3) main steps in GRA such as data normalization, calculating the grey relational coefficient and calculating grey relational grade.

1) Step 1: Data normalization

There are few equations for data normalization in grey relational analysis [6] and the determination of which equations to be employed is based on the characteristic of a data sequences. In this study (1) is employed because the expectancy is the higher-the-better; meaning that the higher GRG represent the more important input factor.

$$x_i^*(k) = \frac{x_i^0(k) - \min x_i^0(k)}{\max x_i^0(k) - \min x_i^0(k)}. \qquad (1)$$

where,

$i = 1,\ldots,m; \quad k = 1,\ldots,n.$

m is the number of experimental data items,

n is the number of parameters,

$x_i^0(k)$ is the original sequence,

$x^*{}_i(k)$ is the sequence after data pre processing,

$\min x_i^0(k)$ and $\max_i^0(k)$ are the smallest and largest value of $x_i^0(k)$.

The range of data is adjusted so as to fall within [0,1] range.

2) Step 2: Calculate the grey relational coefficient.

The second step is to locate the grey relational coefficient by using (2).

$$\xi_i(k) = \frac{\Delta\min + \zeta\Delta\max}{\Delta_{0,i}(k) + \zeta\Delta\max}, \qquad (2)$$

where,

$\xi_i(k)$ = grey relational coefficient at any data point (k),

Δ_{0i} = deviation sequences of the reference sequence and comparability sequence,

$$\Delta_{0,i} = \left\| x_0^*(k) - x_i^*(k) \right\|,$$

$$\Delta\min = \min_{\forall j \in i} \min_{\forall k} \left\| x_0^*(k) - x_i^*(k) \right\|,$$

$$\Delta\max = \max_{\forall j \in i} \max_{\forall k} \left\| x_0^*(k) - x_i^*(k) \right\|,$$

$x_0^*(k) =$ the reference sequence, and

$x_i^*(k) =$ the comparative sequence.

ζ is known as identification coefficient with $\zeta \in [0,1]$. Normally ζ =0.5 is used because it offers moderate distinguishing effect and stability [6,7]

3) Step 3: Calculate the grey realtional grase (GRG)

Finally, to obtain the GRG, the average value of grey relational coefficient is computed.

$$\gamma_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k), \qquad (3)$$

where n is the number of the objective function or the reference sequence, $x_0^*(k)$.

The GRG; γ_i represents the level of correlation between the reference sequence and the comparability sequence. Based on the calculated value of GRG, the grey relational order based on the size of γ_i is constructed. Each γ_i is ordered to the increasing grey relational coefficient. This derived order then gives the priority list in choosing the series that are closely related to the reference series x_o. For example, if $\gamma(x_0, x_i) > \gamma(x_0, x_j)$, then the element x_i is closer to the reference element x_0 than the element x_j. Generally, γ_i > 0.9 indicates a marked influence, γ_i > 0.8 a relatively marked influence, γ_i > 0.7 a noticeable influence and γ_i < 0.6 a negligible influence [8].

B. Backpropagation

The BP algorithm is based on gradient descent or conjugate descent method by calculating the partial derivative of the performance respect to the weight and biases. However, BP suffered from slow convergence and easily trapped in local minima. Below is the BP algorithm adopting stochastic gradient descent method is presented as follows [9]:

- Create a feedforward with n_{in} input, n_{hidden} hidden units, n_{out} output units.
- Initialize all network weights to small random numbers (e.g., between -0.5 and 0.5)
- Until the termination condition is met, DO
- For each pair of the form $\langle \vec{x}, \vec{t} \rangle$, where $\vec{x}$ is the vector of network input values, and $\vec{t}$ is the vector of target network output value, DO

Propagate the input forward through the network:

1. Input the instance $\vec{x}$ to the network and compute the output o_u of every unit u in the network.

Propagate the errors backward through the network:

2. For each network output unit k, calculate its error term δ_k

$$\delta_k \leftarrow o_k(1-o_k)(t_k - o_k) \quad (4)$$

3. For each hidden unit h, calculate its error term δ_h

$$\delta_h \leftarrow o_h(1-o_h) \sum_{k \in outputs} w_{kh}\delta_k \quad (5)$$

4. Update each network weight w_{ji},

where w_{kh} is the weight from unit i to unit j

$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji}, \quad (6)$$

and $\Delta w_{ji} = \alpha \delta_j x_{ji}$, α is the learning rate.

The purpose of network training is to minimize the error function by updating the weight and biases iteratively. The mean square error is normally chosen as error function of ANN training and is written as

$$E_k = \frac{1}{N}\sum_{t=1}^{N}(t_k - o_k)^2 \quad , \quad (7)$$

where t_k is the target at k^{th} pattern and o_k is a networks output pattern at k^{th} pattern and N is the number of ANN pattern.

C. Particle Swarm Optimzation (PSO)

Particle swarm optimization is an evolution computation technique, which is inspired by analogy of the social behavior of birds' flocking and founded by Kennedy and Eberhart [10] in 1995. PSO has a population with random search solution. Each particle is defined as a potential solution to a problem in D-dimensional space. Each candidate solution is referred to as particle. If the optimization problem has D variables, then each particle represents a D-dimensional point in the search space. The fitness of particle is measured using a fitness function that quantifies the distance between the particle and its' optimal solution. Each particle flies through the search space, and its position is adjusted based on its distance from its own personal best position (P_{best}) and the distance from the best particle of the swarm (g_{best}) [7]. The performance of each particle (how close the particles from the global optimum) is measured using fitness function, which depends on the optimization problem. The process is initialized with a group of random particles; m, with particle i representing $\vec{x}_i = [x_{i1}, x_{i2}, ..., x_{id}]$, $i = 1,2,...,m$, the current position of particle. Each particle also contains the current velocity of the particles; $\vec{v}_i = [v_{i1}, v_{i2}, ..., v_{id}]$, P_{best}, the personal best position of particle (cognitive component), and g_{best}, the global best position of the particle swarm (social component). The updates of the particles are accomplished based on the following equations.

$$v_{id}^{(t+1)} = wv_{id}^{(t)} + c_1 rand_1\left(p_{id}^{(t)} - x_{id}^{(t)}\right) + c_2 rand_2\left(p_{gd}^{(t)} - x_{id}^{(t)}\right) \quad (8)$$

$$x_{id}^{(t+1)} = x_{id}^{(t)} + v_{id}^{(t)} \quad (9)$$

Where $p_{id} = p_{best}$, $p_{gd} = g_{best}$, c_1 and c_2 are two positive constants named learning factors and are used to control the movements of particle at each iteration; $rand_1$ and $rand_2$ are two random functions in the range of [0, 1], w is inertia weight, provides balance between the local and global exploration. The value of w is linearly decreases from 0.9 to 0.4 throughout the iteration. While, $v_i(t)$ and $v_i(t+1)$ are current velocity and updated velocity for each of iteration respectively; $x_i(t)$ and $x_i(t+1)$ are current position and updated position for each of iteration respectively.

The advantage of using PSO algorithm over other techniques such as GA; it can be computationally inexpensive, easily implemented, and does not require gradient information of the objective function but only its values. In this study, PSO is applied to the neural network as a training phase to obtain a set of weights and biases that will minimize the error function. Weights are progressively updated until the convergence criterion is satisfied. Here, the error function is the objective function to be minimized by PSO.

III. THE PROPOSED HYBRID MODEL, GRPSO_BP

Fig. 1 demonstrates the architecture of the proposed hybrid-forecasting model. The proposed model comprises of two main modules, a cooperative feature selection (grey relational ANN) and a BP_PSO forecasting engine.

In the first modules, GRA cooperates with ANN to identify the most significant affecting factors. Then it is followed by the number of optimum input nodes for BP_PSO forecaster. The combination of GRA and ANN is called as cooperative feature selection since both of them are cooperating with each other in identifying the optimum significant affecting factors of KLSE closing price. The task of the second module is to find the best-forecasting model for predicting KLSE closing price based on the input nodes suggest by Grey relational ANN.

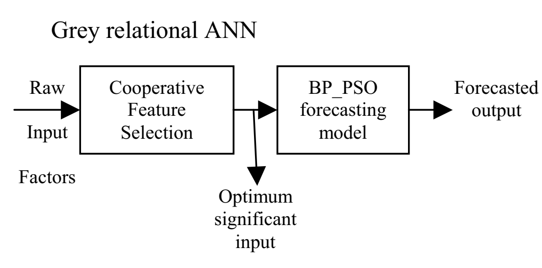

Figure 1: The architecture of the proposed model: GRPSO_BP

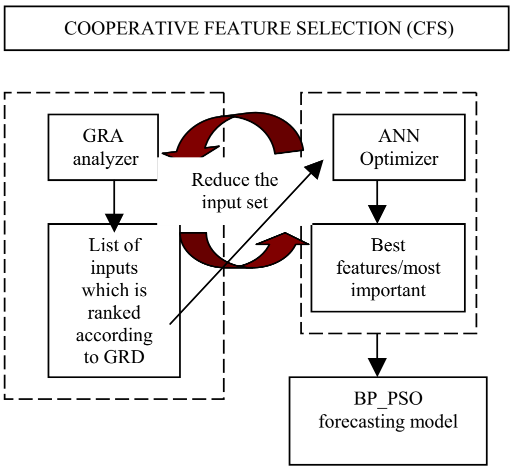

Figure 2: Cooperative Feature Selection Components

The following sections will discuss in detail the above modules.

A. Cooperative Feature Selection (CFS): Grey relational ANN

Fig. 2 illustrates the Cooperative Feature Selection (CFS). Two phases are involved in CFS: GRA analyzer and ANN optimizer. The procedures are given as below.

1) GRA analyzer: GRA analyzer is used as preprocessing step and it is self-governing. GRA analyzer will examine the relevancy level between each predictor and the dependent variables. Finally, GRA will rank each predictor according to its importance or priority. Fig. 3 shows the ranking scheme that contains the 14 affecting factors for KLSE, which are priorly analyzed by GRA. This ranking scheme is ordered in descending order based on the GRG values. The highest GRG value is ranked at the first ordered (Composite index) followed by the smaller GRG values. The smallest GRG value implied the least important of affecting factors, in this case the property index. Then, the affecting factors that have GRG less than 0.6 is deleted from the ranking scheme [8]. Therefore, Technology index, Construction Index, Mining Index and Property Index are excluded from the ranking scheme; only nine input left to be trained further by ANN optimizer.

2) Neural network optimizer: Selected output from GRA analyzer will be the input for ANN optimizer. There are three iterative processes involved; testing the ANN accuracy, judging the data and deducing from the data input factors set. Here, the input factor that produces the minimum drop off in predictive ANN performance is eliminated. The output from this phase is the optimum number of predictors or optimum input factors, which is ranked based on their precedence. These optimum input factors are utilized as input nodes to build the BP_PSO forecaster. From Fig. 3, instead of using 14 affecting factors, only four affecting factors (Composite,

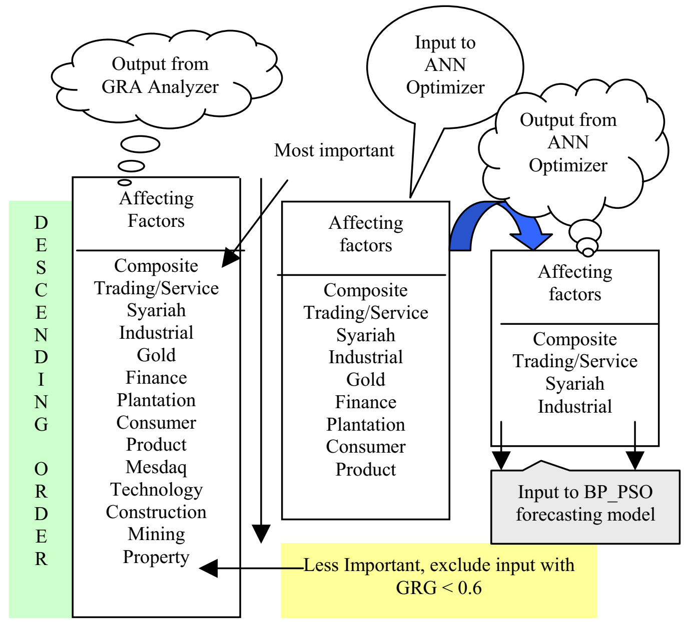

Figure 3: Implementation of Cooperative feature selection

Trading/Service, Syariah and Industrial) or ($SI, TSI, CmpI$ and II) are sufficient for predicting the next day KLSE closing price.

B. PSO-BP forecasting model

Fig. 4 below illustrates the two main part of BP_PSO forecasting model; PSO model and BP model. Here, four input factors that yield from CFS phase are used as input nodes to PSO model in the process of finding the initial weights and threshold for BP neural networks. Then these initial weight and threshold are fed up into ANN model with fixed network architecture, and then ANN model is trained with BP algorithm. Here, the structure of ANN used is 4-9-1, where 4 represent four input nodes, 9 represent nine hidden nodes and 1 represent one output node. The learning rate and momentum used to train ANN are [0.9, 0.5]. The sigmoid function is used as transfer function. The learning error ε is set as 0.05.

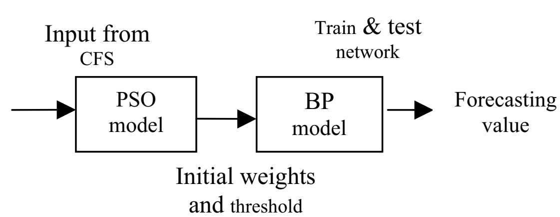

Figure 4: Component of BP_PSO forecasting model

Meanwhile, for PSO: swarm scale m =20, initial weight, w decreases linearly from 0.9 to 0.4 along with iteration, c_1 = 1.5 and c_2 = 1.0, $V_{\max}$ =1.15 and $G_{\max}$ (iteration) = 100

IV. RESULTS AND DISCUSSION

A. Ex;erimental Data and Performance Measures

The performance validation of the proposed model, GRBP_PSO is conducted on KLSE closing price. KLSE data contains 200 observations of daily Kuala Lumpur Stock Exchange (KLSE) closing price from 4th January 2005 till 21st October 2005 and 14 affecting factors that influence te movement of KLSE closing price. These affecting factors are consumer Index (CI), construction index (CoI), gold index (GI), finance index (FI), product index (PI), Mesdaq index (MI), mining index $(MinI)$, plantation index (PlI), property index $(\Pr oI)$ (ProI), syarian index (SI), technology index (TI), trading/ service index (TSI), composite index $(CptI)$ and industrial index (II).

These data are split into training, validation and testing. Data from 4th Jan, 2009 till 9h Sep, 2005 are used as training data set and are used exclusively for model development. The data from 4th Jan, 2009 till 30th Sept, 2009 are employed to validate the performance of model generalization. In order to evaluate the model capability, the proposed model is tested on the unseen data that is never been used in training and validation phase; 3rd October, 2005 till 20th October, 2005. The test data should not be used in the model estimation or model selection process to ensure that the real forecasting performance is evaluated. To measure the performance of GRBP_PSO, four statistical tests are used namely, mean square error (MSE), root mean square error (RMSE), mean absolute deviation (MAD) and mean absolute percentage error (MAPE).

B. Results

Among 14 factors initially used in this study, only four factors are selected by Grey relational ANN (CFS) as dominant factors that influenced the KLSE closing price. The factors include:Composite index ($CmpI$), Trading/Service index (TSI), Syariah index (SI) and Industrial index (II). This information are beneficial to people who are directly involved in stock market trading. These four indexes can be used as indicator for KLSE closing price. The traders can predict the price of the KLSE closing price based on the values of these four indexes. For example, if these indexes are high they may expect the closing price is also high or otherwise. Therefore, this information can assist them in determining the appropriate time to sell or to buy the stocks to maximize their profit.

Meanwhile, Table 1 shows the result obtained from GRBP_PSO, BP_PSO and BP. Table 1 shown that the result obtained from BP model alone without PSO is the worst compared to BP_PSO and GRBP_PSO.

TABLE 1: Statistical test for each model

Statistical test	GRBP_PSO	BP_PSO	BP
MSE	3.37	604.07	1290.25
RMSE	1.84	24.58	35.92
MAD	1.78	24.17	35.90
MAPE	0.19	2.69	3.88
Accuracy	98%	96%	94%

This indicates that the application of PSO to determine the initial values of BP weights and thresholds has improved the BP convergence and accuracy performance. From Table 1, it was obviously demonstrated that the performance of the proposed model, GRBP_PSO is better than BP_PSO. The MSE, RMSE, MAD, and MAPE of GRBP_PSO were lower than BP_PSO. This indicates that the combination of GRA and ANN has removed the irrelevant factors that cause the instability in the BP_PSO forecasting model that lead to the poor forecasting performance. Besides that, the application of GRA also has successfully simplified the network structures; reduced 73% of the input number to be used in BP_PSO. Consequently, this simplified network structures shortened the training time to train the network to converge to the optimum solution.

Fig. 5 shows the forecasting values produce by each forecasting model. From Fig. 5, it clearly shown that the forecasted values produce by GRBP_PSO is more closely to the actual values. This indicates that the application of GRA and PSO can help the BP network to automatically identify the optimum affecting factors and can minimize the probability of BP network stuck in local minima. Therefore by using the proposed model, the dependency on human expert domain knowledge in determining the optimum affecting factors is no longer needed. Moreover, the forecasting accuracy of the proposed model also is the highest (98%) compared to BP_PSO and BP.

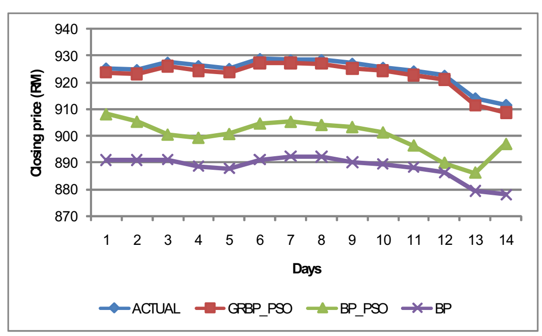

Figure 5: Comparative performance forecasted values and actual data

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
GRBP_PSO	14	922.5343	5.5696	1.4885
BP_PSO	14	900.1445	6.1641	1.6474

Figure 6(a): Statistic for GRBP_PSO and BP_PSO

One-Sample Test

	Test Value = 924.31					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
GRBP_PSO	-1.193	13	.254	-1.7757	-4.9915	1.4401
BP_PSO	-14.669	13	.000	-24.1655	-27.7245	-20.6064

Figure 6(b): T-test for GRBP_PSO and BP_PSO

To further validate the performance of the proposed model, the significance test is conducted. Fig. 6(a) shows the statistic values for GRBP_PSO and BP_PSO. It shows that the mean of GRBP_PSO (922.53) is closer to the mean of the actual values (924.31) than the mean of BP_PSO (900.14). The standard mean error and the standard deviation of GRBP_PSO are lower than BP_PSO. Fig. 6(b) shows the t-test results using test values =924.31 for GRBP_PSO and BP_PSO. The purpose of this test is to investigate whether there is statistical significant difference between the mean of actual values (μ_{actual}) and the mean of GRBP_PSO (μ_{GRBP_PSO}) or BP_PSO (μ_{BP_PSO}).

From Fig. 6(b), it shows that the P > 0.05 for GRBP_PSO (refer column Sig. 2 tailed). Therefore, the H_{null} is accepted. Furthermore, the difference between the upper and lower value for 95% interval are range between –4.9915 and 1.4401, and it indicates that there is no statistically difference between the mean of actual data and the mean obtained from predicted value of GRBP_PSO. Therefore, we can conclude that there is no prediction accuracy difference in actual data and GRBP_PSO. This result indicates that the predicted values given by GRBP_PSO can represent the actual values of KLSE closing price.

However, the contrast result obtained from BP_PSO where the P < 0.05 and the H_{null} is rejected. This indicates that there is a difference between the mean of the actual value and the mean of BP_PSO. Furthermore, the upper and the lower range of 95% confidences are both in negative. This indicates that $\left(\mu_{actual} - \mu_{BP\_PSO} < 0\right)$ and $\mu_{actual} < \mu_{BP\_PSO}$. This result implies that the forecasting error gained from BP_PSO is high therefore it is not suitable to be selected to predict the KLSE closing price. Thus, from the experimental result and significance test, it showed that the proposed model GRBP_PSO is able to give better result than BP_PSO and BP.

V. CONCLUSION.

In this paper, hybridization of GRA with ANN was adopted as cooperative feature selection to find and to identify the optimum affecting factors that affect the KLSE closing price. To overcome the local minima problem in BP, PSO was employed to determine the initial weights and thresholds of BP networks. Thus, with better learning ability and generalization ability, GRBP_PSO model is relatively stable and can converge faster globally. To evaluate the performance of GRBP_PSO model, the comparison with BP_PSO and BP are implemented. Experimental results show that GRBP_PSO model yields lower forecasting error compared to both other models.

ACKNOWLEDGMENT

Authors would like to thank Research Management Centre (RMC) Universiti Teknologi Malaysia, for the research activities, *Soft Computing Research Group* (SCRG) and Dr Suarni Sumormo (MRB) for the support in making this study a success.

REFERENCES

[1] A.S.Chen, M.T.Leung, & H. Daouk.," Application Of Neural Networks To An Emerging Financial Market: Forecasting And Trading The Taiwan Stock Index". Computers And Operation Research, 30, 901-923,2003.

[2] K.Mehdi, H. R. Seyed, B. Mehdi, "A new hybrid ANN and fuzzy regression model for time series forecasting". Fuzzy sets and systems, 159 ,769-786,2008.

[3] S.M.Shamsuddin, M.N.Sulaiman, M.Darus." An Improved Error signal for the BP model for Classification Problems", Int. Journal Of Computer Mathematics, 76:3,297-305.2007.

[4] J.L.Deng. "Control Problems Of Grey Systems". Systems And Control Letters, Vol.5, pp.288-294,1982.

[5] J.L.Deng," Introduction To Grey System Theory", The Journal Of Grey System, Vol.1. pp,1-2, 1989.

[6] R.Sallehuddin, S.M.Shamsuddin, S.Z. Hashim, A. Abraham, "Forecasting Time Series Data Using Hybrid Grey Relational Artificial Neural Network and Auto Regressive Integrated Moving Average", Neural Network World 6/07,573-605,2007.

[7] S.J.Lin,I.J. Lu, L.Lewis, "Grey Relation Performance Correlations Among Economics, Energy Use and Carbon Dioxide Emission in Taiwan", Energy Policy 35, 1948-1955 ,2007.

[8] C.Fu, J. Zheng, J. Zhao, and W.Xu, "Application of Grey Relational Analysis for Corrosion Failure of Oil Tubes", Corrosion Science 43, 881-889,2001.

[9] J. Kennedy, and R.C..Eberhart.."Particle Swarm optimization". 1995. Proceedings of IEEE International Conference on Neural Networks, Piscataway,NJ, p:1942-1948,1995.

[10] Y. Shi, R.C. Eberhert.: Parameter Selection in Particle Swarm Optimization. Evolutionary Prgramming, VII, Proc.EP 98, 591-600, 1998