A HYBRID MODEL COMBINING WITH BEST REPLACEMENT OPTIMIZATION TECHNIQUES AND ROUGH SET THEORY FOR STOCK MARKET PREDICTION

Dr. T.Chitra kalarani* and S.Indrakala**

*Associate professor, Kundavai Nachiaar Govt. Arts College(W) Autonomous, Thanjavur.

**Assistant professor, Kundavai Nachiaar Govt. Arts College(W) Autonomous, Thanjavur.

ABSTRACT

Stock price prediction is an important task for practitioners and professional analysts. However, it is a tough problem because of the uncertainties involved in prices. It is all the times true shareholders in general get loss because of uncertain investment purposes and unsighted assets. In this paper, we propose and implement a hybrid model combining Best Replacement Optimization Techniques (BSO) and Rough Set Theory (RST) to find optimal buy and sell of a share on Nifty fifty Stock Index . Our experimental results exhibit that the recommended hybrid model has higher precision than other considered forecasting models selected for this study. We believe that findings of this paper will be useful for stock investors to make a decision about optimal buy and/or sell time on this stock market.

1. Introduction

Forecast of stock prices has been regarded as one of the most challenging applications of modern time series forecasting. Thus, numerous models have been depicted to provide the investors with more precise predictions. Hybrid forecast is a well-established and well-tested approach for improving the forecasting accuracy. Therefore, the importance of hybrid forecast methods has steadily increased and it acts still on time series forecasting. In order to enhance the forecasting performance of the time series models, the hybrid model combined by the time series models and other models have been advanced many researches. In recent years, more hybrid forecasting models have been proposed. For example, Auto-regressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) are applied to time series forecasting as it reduces model uncertainty which typically occurs in statistical inference and time series forecasting. Pai and Lin [3] proposed a hybrid methodology to exploit the unique strength of ARIMA models and support vector machines (SVMs) for stock prices forecasting. Armano, et al. [1] presented a new hybrid approach that integrated artificial neural network with genetic algorithms (GAs) to stock market forecast. Tsaih et al. [6] presented a hybrid artificial intelligence (AI) approach that integrated the rule-based systems technique and neural networks to S&P 500 stock index prediction. Hyun-jung Kim , Kyung-shik Shin [4] proposed the effectiveness of a hybrid approach based on the adaptive time delay neural networks (ATNNs) and the time delay neural networks (TDNNs) with the genetic algorithms (GAs) in detecting temporal patterns for stock market prediction tasks. Ching-Hseue cheng, Tai-Liang chen, Liang-Ying Wei in [3] this paper proposed a hybrid forecasting model using multi-technical indicators to predict stock price trends. Kai Keng Ang, Chai Quek [5] used neuro-fuzzy systems and neural networks for forecasting stock price difference on artificially generated price series data.

To improve upon past forecasting models, a revised model should be able to overcome the drawbacks contained in previous models and should offer a good methodology which could be used more easily by investors. Empirically, this paper employs two types of stock databases (stock index and individual stock price) as experimental datasets. From the model verification, it is shown that the refined processes are effective in improving forecasting accuracy, and, based on the evidence, a stock analyst or investor can employ the refined processes proposed in this paper to improve their forecasting tools or models.

The rest of this paper is organized, as follows: Section 2 demonstrates the proposed model and algorithm; Section 3 evaluates the performance of the proposed model and describes the findings; and Section 4 draws conclusions and proposes recommendations for future research.

2. PROPOSED FORCASTING MODEL

The Best Replacement Optimization Model with Rough Set (abbreviated as BRO-RS) is a hybrid complex data prediction algorithm is used as a best solution to solve complex stock market data in future prediction. The Hybrid Dynamic algorithm BRO-RS uses it particle to detect best fitness and Rough set to Reduce the dimension of reduct sets in Stock Market data, The BRO does not require any gradient information of the function to be optimized and uses only primitive mathematical operators. Rough set theory offers a novel approach to manage uncertainty that has been used for the discovery of data dependencies, importance of features, patterns in sample data, feature space dimensionality reduction and the classification of objects. While rough set on their own provide a powerful technique, it is often combined with other computational intelligence techniques such as neural networks, fuzzy sets, genetic algorithms, Bayesian approaches, swarm optimization and support vector machines. BRO as a new evolutionary computation technique, in which each potential solution is seen as a particle with a certain velocity flying through the problem space. The Particles find optimal regions of the complex search space through the interaction of individuals in the population.

BRO is attractive for feature selection in that particle swarms will discover best feature combinations as they fly within the subset space. Compared with other evolutionary techniques, BRO requires only primitive and simple mathematical operators. In this research, Rough set is applied to improve feature selection and data reduction. BRO is used to optimize the rough set feature reduction to effectively classify stock prices. The input Dataset is processed by using Rough set, the rough set helps to detect reduct dataset. In Reduct sets the Swarm is used to assign a particle. It works all particles tend to fly towards better and better positions over the searching process until the swarm move to close to an optimum of the fitness function. As compared with other optimization methods, it is faster, cheaper and more efficient. In addition, there are few parameters to adjust in BRO. That's why BRO is an ideal optimization problem solver in optimization problems. BRO is well suited to solve the non-linear, non-convex, continuous, discrete, integer variable type problems. BSO is initialized with a population of random solutions. The advantages BRO is easy to implement and there are few parameters to adjust.

In BRO each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. This value is called $pbest$. Another "best" value that is tracked by the global version of the particle swarm optimizer is the overall best value, and its location, obtained so far by any particle in the population. This location is called $gbest$.The particle swarm optimization [29] concept consists of, at each time step, changing the velocity each particle toward its $pbest$ and $gbest$ locations as follows. A gene with highest fitness value is taken and all possible combinations of the selected gene with the other genes are constructed. Fitness of the selected genes with different combinations is calculated. If the current particle's fitness evaluation is better than the $pbest$, then this particle becomes the current best and its position and fitness are stored. Then, the current particle's fitness is compared with population's overall previous best fitness. If the current value is better than gbest, then this is set to the current particle's position, with the global best fitness updated. The velocity and position of the particle is then updated. This process is carried out until the stopping criterion is met, usually a maximum number of iterations. In BRO with Rough Set the process iteration is limited with the best Fitness threshold value

ALGORITHMS

In BRO the $Gbest$ parameter is adjusted by limiting the iteration process by efficient fitness Procedures as Follows:

Best Replacement model Using Rough Set:

Input: Stock Market Data High Price, Low Price from Nifty Fifty ***Companies*** Where Set
$\boldsymbol{S = \{S_1 U\, S_2\, U\ \ldots .\ S_{50}\}}$ **(Union of 50 Subsets)**

Output: Fitness Function Based on Reduct Set of Rules

Begin

Swarms $\{x_{id,}v_{id}\} = Generate\ (m)$: /* Initialize a population of particles

with random positions and velocities on S dimensions*/

$Pbest\ (i) = 0;\ i\ = 1, \ldots, m, d\ = 1, \ldots, S$

$Gbest = 0;\ Iter = 0;$

$For\ each\ Subset\ of\ Set\ S\ identify\ the\ reduct\ sets\ by\ using\ Algorithm 1$

$In\ R \rightarrow Reduct\ Set\ (S)$

$\{For\ (each\ particle\ in\ R)$

$\{While (Iter < MaxGen\ and\ Gbest < MaxFit)$

$\{For(every\ particle\ i)$

$\{Fitness(i) = Evaluate(i);$

$IF(Fitness(i) > Pbest(i))$

$\{Pbest(i) = Fitness(i);\ p_{id} = x_{id}\ ;\ d = 1, \ldots, S\}$

$IF(Fitness(i) > Gbest(i))$

$\{Gbest = Fitness(i);\ gbest = i;\ \}$

$For(every\ particle\ i)$

$\{For(every\ d)$

$\{v_{id} = w * v_{id} + c_1 * R_1 * (p_{id} - x_{id}) + c_2 * R_2 * (p_{gd} - x_{id})$

$IF(v_{id} > MaxV)\{v_{id} = MaxV;\}$

$IF(v_{id} < -MaxV)\{v_{id} = -MaxV;\}$

$x_{id} = x_{id} + v_{id}\}$

}

$Iter = Iter + 1;$

}

/* R_1 and R_2 are two Constant Variables in the range [0,1]*/

$Detect\{gbest\}$

$While\ (GF - 1)$

{

$v_{id} = w * v_{id} + c_1 * R_1 * (p_{id} - x_{id}) + c_2 * R2 * (p_{gd} - x_{id})\}$

do

}

$GF = GF + 1$

}

End

}

PSEDOCODE: BSO WITH ROUGH SET

Step1: The Input dataset S is taken from Nifty fifty Companies,

Step2: The dataset S is used by roughset to detect reduct sets, the reduct sets is initialize as particle.

Step3: Initialize position and velocity of all the particles randomly in the N dimension space.

Step2: Evaluate the fitness value of each particle, and update the global optimum position.

Step3: According to changing of the gathering degree and the steady degree of particle swarm, determine whether all the particles are re-initialized or not.

Step4: Determine the individual best fitness value. Compare thel_p of every individual with its current fitness value. If the current fitness value is better, assign the current fitness value to l_p.

Step5: Determine the current best fitness value in the entire population. If the current best fitness value is better than the g_p , assign the current best fitness value to g_p .

Step6: For each particle, update particle velocity,

Step7: Repeat the iteration of the particle using $gbest$ fitness value and limit the Iteration of the particle.

Step8: Update particle position.

Step9: Repeat Step2 - 7 until a stop criterion is satisfied or a predefined number of iterations are completed. While maximum iterations or minimum error criteria is not attained Particles' velocities on each dimension are clamped to a maximum velocityv_{max}. If the sum of accelerations would cause the velocity on that dimension to exceedv_{max}, which is a parameter specified by the user. Then the velocity on that dimension is limited to v_{max}.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

For evaluating the performance of the proposed BRO with Rough set forecasting model, the daily ICICI Company price is used in this study. In forecasting, the five technical indicators, namely the previous day's cash market high (x_1), low (x_2), volume (x3), today's opening cash index (x4), and 10-days total amount weight stock price index (TAPI 10) (x5), determined by the review of domain experts and literatures (Wood, 2002; Leigh et al., 2005) are selected as the forecasting variables for predicting TAIEX closing cash index. Please refer to Wood (2002) and Leigh et al. (2005) for more details about technical indicators. The five financial time series datasets and the daily TAIEX closing cash prices in the TAIEX dataset are depicted in Figure 3 and 4, respectively. The daily data of technical indicators and cash prices from January 2, 2003 to February 27, 2006 of TAIEX cash index provided by Capital Futures Corporation, Taipei, are collected. There are totally 781 data points in the dataset. Among, the first 623 data points (79.77% of the total sample points) are used as the training sample while the remaining 158 data points (20.23% of the total sample points) are used as the testing sample

BRO with Rough set require additional parameter settings for their operation. These are given in Table 1.

Population	3145
Generation	251
C1	2
C2	2
Weight	1.4~0.4

Table 1 BSO with Rough Set parameter settings

In BSO with Rough Set, the inertia weight decreases along with the iterations, varying from 1.4 to 0.4 according to the equation (1).

$$Weight = (weight - 0.4) * (MAXITER - Iter) / MAXITER + 0.4 \quad (1)$$

Date	Open Price	High Price	Low Price	Close Price	Total Volume	No of Trades
09/11/14	179.1	174.2	177.3	175.2	56,91,537	49,335
09/10/14	177.85	171.7	172.95	177	63,45,483	73,091
09/09/14	176.7	172.75	176.05	174.15	53,81,962	40,668
09/08/14	178.5	170.1	176.5	176.8	89,29,544	67,188
09/05/14	176.65	164.7	166	175.45	1,72,40,111	1,18,797
09/04/14	176	156.2	175.2	167.5	2,02,51,203	1,45,223
09/03/14	186.9	180.5	185.65	182.9	75,97,803	53,178
09/02/14	185.3	178.85	179.65	184.05	1,00,55,042	68,222
09/01/14	180.6	175.45	178	179.75	98,33,749	67,952
08/28/14	183.35	176	181.1	177.3	1,45,58,542	1,04,275

Where MAXITER is the maximum iteration (generation) and Iteration is the current iteration.

Table 2 indicates Sample Testing dataset of ICICI Company.

Parameters	High Prices		Low Prices	
Dataset	Analysis Set(Training)	Validation Set(Test)	Analysis Set(Training)	Validation Set(Test)
Total Objects	143191	3145	143191	3145
Objects Covered	3145	251	3145	251
Min Support	105.56	128.19	105.56	128.19
Max Support	2572	1170.25	2572	1170.25
Average Support	576.4903666	270.6402703	575.993815	270.6402703
Min Accuracy	0.024127521	0.001036881	0.001367952	0.002429055
Max Accuracy	0.025186772	0.002719556	0.003187446	0.003731343
Average Accuracy	0.024900086	0.002166111	0.00189906	0.002930401

Table 3 Statistical Analysis of Input Training Dataset of ICICI Company.

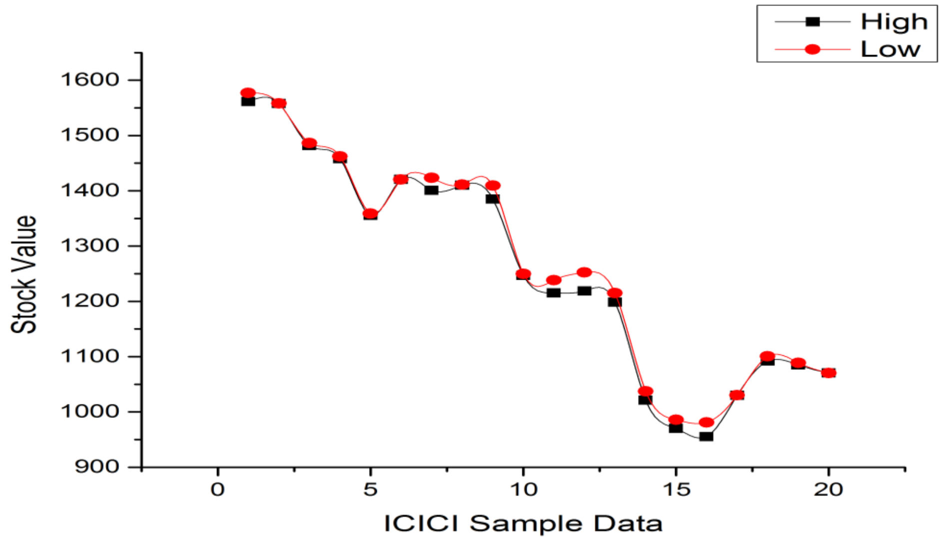

Figure 1 BRO with Rough Set Prediction of Low Price Value and High Price Value in ICICI Company.

The Figure 1 indicates the BRO with Rough Set Prediction of Low Value and High Value in Stock Market Data from Jan 2008 to Sep 2014. The Sample Data is used to detect GP and predict the low value and high value for ICICI Company.

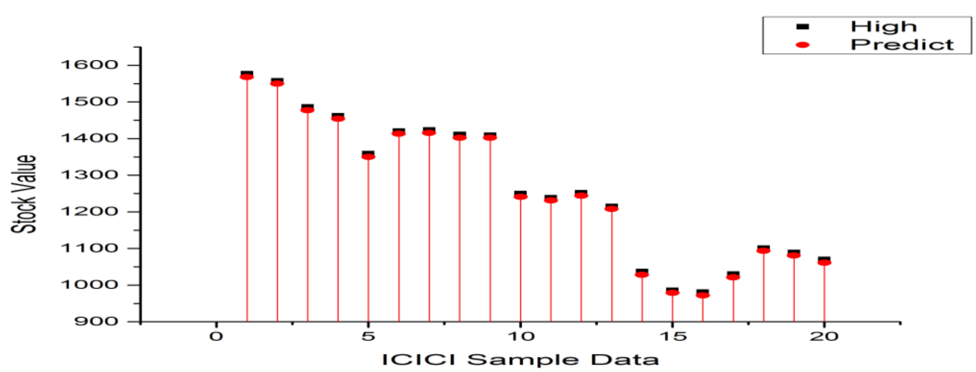

Figure 8.2 BRO with Rough Set Prediction of High Price Value in ICICI Company

The Figure 8.2 indicates the BRO with Rough Set Prediction of High Value in Stock Market Data from Jan 2008 to Sep 2014. The Sample Data is used to detect GP and predict the high value for ICICI Company.

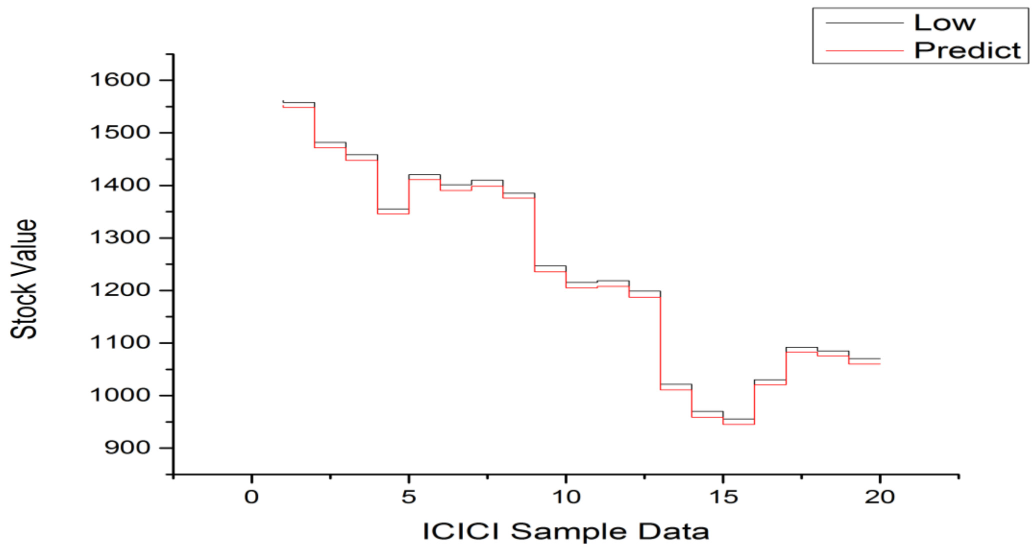

Figure 8.3 BRO with Rough Set Prediction of Low Price Value in ICICI Company

The Graph 8.3 indicates the BRO with Rough Set Prediction of Low Value in Stock Market Data from Jan 2008 to Sep 2014. The Sample Data is used to detect GP and predict the Low value for ICICI Company.

8.5 CONCLUSION

In this study we have investigated the basic aspect of the prediction problem of a stock market. In this chapter we have only used the Historic prices of the Index values for prediction. A new hybrid dynamic BRO and RS model is proposed to increase the prediction accuracy for both short term and long term stock market indices prediction. This new algorithm uses Rough set to detect reduct sets and enables all particles in the reduct sets using swarm to perform the global search using effective *gbest* fitness in the whole search space. The proposed model is obviously better than the standard existing model in processing the reduct set as particles, where in existing method, the particles are subjected to normal operators are capable in performing the global search. In a highly volatile market like Indian Stock Market, the performance levels of the BRO with RS, reported in the paper will be very useful. Especially, the prediction of the direction of the market with fairly high accuracy, will guide the investors and the regulators. In Future work, the proposed dynamic algorithm is used in different area of fields depending time series data for future prediction.

References:

[1]. Armano G., et al.,(2005), "A hybrid genetic-neural architecture for stock indexes forecasting", Information sciences, volume 170, page(s): 3-33, 2005.

[2]. Ching-Hsue cheng, Tai-Liang Chen, LiangYing Wei (2010), " A hybrid model based on rough set theory and genetic algorithms for stock price forecasting", Page(s): 1610-1629.

[3] Pai P.F., Lin C.S. (2005), A hybrid ARIMA and support vector machines model in stock price forecasting; Omega Volume 33; P)age(s):497–505.

[4]. Hyun-jung Kim , Kyung-shik Shin,(2007) "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets" , Applied Soft Computing, Volume 7 , Page(s): 569–576.

[5]. Kai Keng Ang, Chai Quek,(2006) "Stock Trading Using RSPOP: A Novel Rough Set-Based NeuroFuzzy Approach" , IEEE TRANSACTIONS ON NEURAL NETWORKS.

[6]. Tsaih R, Hsu Y, Lai CC. ,(1998), "Forecasting S&P 500 stock index futures with a hybrid AI system". Decision Support Systems Volume 23,Issue 2, Page(s): 161–174.