National Conference on Emerging Technologies for Sustainable Engineering & Management (NCETSEM'18)-2018

RESEARCH ARTICLE OPEN ACCESS

Stock Selection using Differential Evolution Algorithm

Keerthana G[1]

[1]ME Software Engineering & Department of Computer Science and Engineering & College of Engineering Guindy, Anna University, Chennai (India)

[1]meenaloshini14@gmail.com

ABSTRACT:

Stock selection is a method for selecting a stock for investment. The stock investment can be long or short, depending on the investor's expectation. Depend upon the price movement of the stock, investors can buy a stock or sell a stock. Stock selection is one of the most crucial but challenging issues, due to the time complexity of financial market. Genetic Algorithm outperforms in Stock market prediction but it has some drawbacks, such as time-consuming and constructs more generation for finding optimized weight. In order to overcome the Genetic Algorithm problem, Differential Evolution (DE) algorithm is used. DE has two advantages; fast convergence and using few control parameters. Therefore, DE algorithm seems to be promising approach for weight optimization. In the proposed work, technical features are constructed using fundamental features. Technical features are used for finding the price movements of the stock. The technical features are normalized using Z-Score Normalization and their feature weight has been computed using Differential Evolution Algorithm. Feature weight and normalized score are linearly combined to compute the stock score. Based on the stock score, stocks are ranked and it has been used for the prediction of next year stock.

Keywords **— Stock Selection, Differential Evolution, Z-Score Normalization, Prediction.**

.

I. Introduction

Stock selection is a method for selecting a stock for investment. The stock investment can be long or short, depending on the investor's expectation. Depend upon the price moving of the stock, investors can buy a stock or sell a stock. Stock selection is one of the most crucial but challenging issues, due to the complexity of financial markets. So it remained a challenge for the common investors, stock buyers/sellers, policy makers, market researchers and capital market role players to gain knowledge about the daily stock market price values.

Making money and gaining high profit is the dream of every investor, but it requires proper financial knowledge with the particular stock market data. A lot of risks are involved in the stock market; the investors become highly insecure to invest their amount. Though a lot of research has already been done on US Stock market, European Stock market, China, and Japan Stock markets but less attention are given to Indian Stock market in comparison to them. Due to rapid growth in IT, Telecom sectors in the last decade Indian Stock market is setting its upcoming platform not only in Asia as well as globally. Asia's oldest stock exchange i.e. Bombay Stock Exchange (BSE) and Nifty stock Exchange (NSE) is contributing a lot to the global economy. It has been found that not only the economic factors but also individual investors.

In existing literature, Stock market analysts have adopted two categories: traditional statistical regression approaches and computational intelligence (CI) techniques. Both of them have their respective strength and weakness. Traditional statistical regression models are relatively easy to implement and understand due to their simple forms; nevertheless, they often appear relatively poor performance. Due to the time complexity in stock markets, the CI models have been more efficient than the traditional statistical models, though might be difficult to understand. CI models have been applied to stock evaluation, such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and various optimization algorithm (e.g., Genetic Algorithm (GA)).

In proposed work, technical features are constructed using fundamental features. Technical features are used for finding the price movements of the stock. The technical features are normalized using z-score normalization and the technical features used for calculating feature weight. Feature weight calculation computes using Differential Evolution Algorithm. Feature weight and normalized score are linearly combined to form a stock score. Depend upon the stock score, it can be ranked and the top-ranked stock has higher potential to raise its price in the future. According to their mean calculation, next year can be estimated as an average return of the stocks.

II. SYSTEM DESIGN

The proposed system is designed using Nifty Stock Exchange (NDE) dataset. The dataset contains fundamental features such as date, open price, high price, low price, close price and volume. Open, high, low and close price describes the price movement of the stock on a particular day. Volume is the total stock to sell and buy on a particular day. Using fundamental features, technical features are constructed. The technical features are normalized using z-score normalization and their feature weights are calculated using Differential Evolution Algorithm. The algorithm contains 5 steps namely population initialization, fitness function, mutation, crossover, and selection. Feature weight and normalized score are linearly combined to calculate the stock score. Based on the stock score, stocks are ranked and the ranked stocks are estimated for the prediction of next year stock.

A. System Architecture

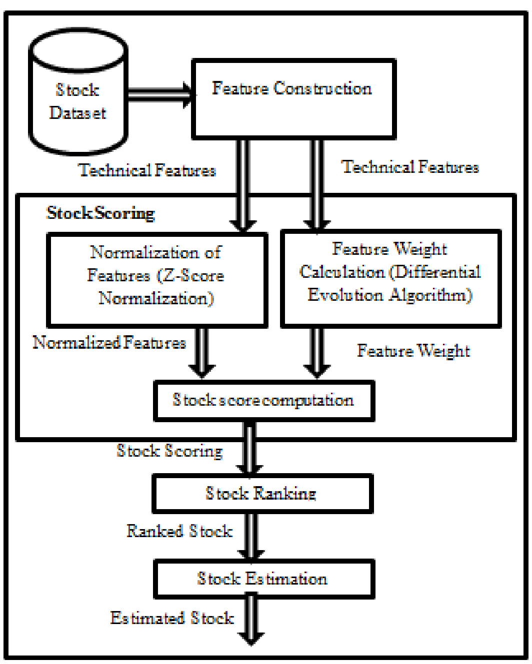

Fig 1: System Architecture

The stock dataset is an input to the system. The dataset contains the fundamental feature such as date, open, close, low, high and volume. Open, high, low and close price describes the price movement of the stock on a particular day. Volume is the total stock to sell and buy on a particular day. Using the fundamental features technical features are constructed. The technical features are Average True Range (ATR), Relative Strength Index (RSI), Rate of Change (ROC), Return on Assets (ROA), Simple Moving Average (SMA), Simple Moving Average Difference (SMADIFF), Force Index (FI) and Williams %R. The technical features are normalized using z-score normalization and their feature weights are calculated using Differential Evolution Algorithm. The algorithm contains 5 steps namely population initialization, fitness function, mutation, crossover, and selection. In population initialization, parameters are initialized. Mutation and crossover process find all possible new generation values. Finally, selection process finds the best-optimized feature weight. Feature weight and normalized score are linearly combined to calculate the stock score based on the stock score, stocks are ranked. The ranked stocks are estimated to predict the next year stock.

B. Feature Construction

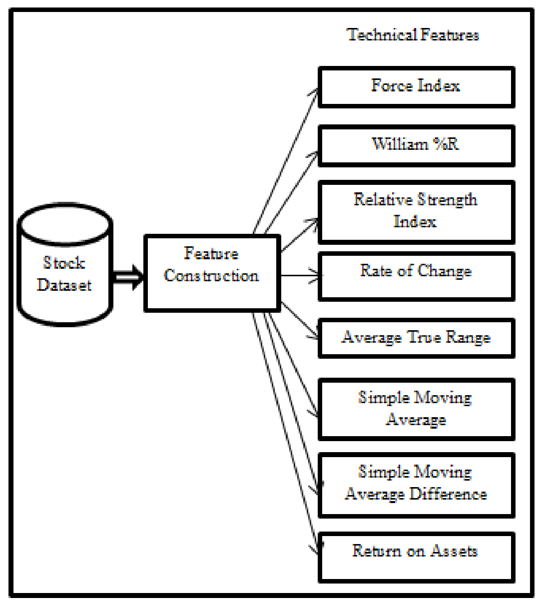

Fig 2: Feature Construction

Feature construction attempts to increase the expressive power of the original features usually the dimensionality of the new feature set is expanded and it is bigger than original feature set. Assuming n features $A_1, A_2 \ldots A_n$ after feature construction it have an additional m features $A_{n+1}, A_{n+2} \ldots A_{n+m}$. The technical features which includes Force Index, William R%, Relative Strength Index, Rate of Change, Average True Range, Simple Moving Average, Simple Moving Average Difference and Return on Assets. Return on Assets is a percentage of profit that company earns in relation to its overall resources. The Rate of Change, which refers to a simple momentum that measures the percentage change in price from one period to the next. The

Relative Strength Index is a momentum indicator that compares the gains and losses over a specified time period to measure speed and change of price movements.

The Williams %R is a momentum indicator that measures overbought and oversold levels. It compares the close price of a stock to the high price and low price over a period of time. The Average True Range is used in technical analysis to measure volatility. A Simple Moving Average is an arithmetic moving average calculated by adding the closing price and dividing the total by the number of time periods. The Force Index is an oscillator that fluctuates above and below zero. It combines price movement and volume to assess the force behind price movements.

C. Feature Weight Calculation

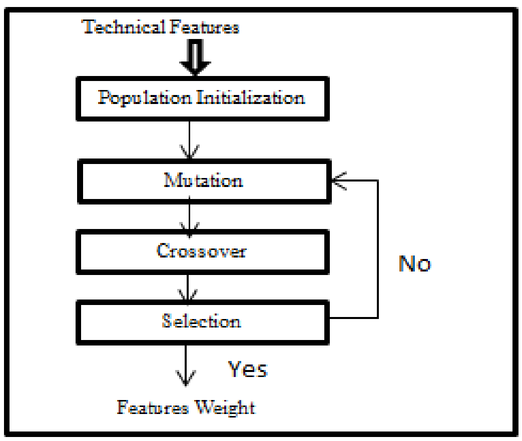

Fig 3: Feature Weight Calculation

Feature weight calculation using Differential Evolution Algorithm. DE algorithm undergoes following five steps. They are population initialization, fitness function, mutation, crossover, and selection process. In population initialization, parameters are fixed. Mutation process which randomly chooses the three values as parent value and mutates them and find the possible values. Crossover process which chooses the possible mutated values and crossover them and finds the new values. Finally, a selection process which checks the fitness value and finds whether the process goes into the loop or generate the next value as the weight of the feature.

D. Stock Scoring And Ranking

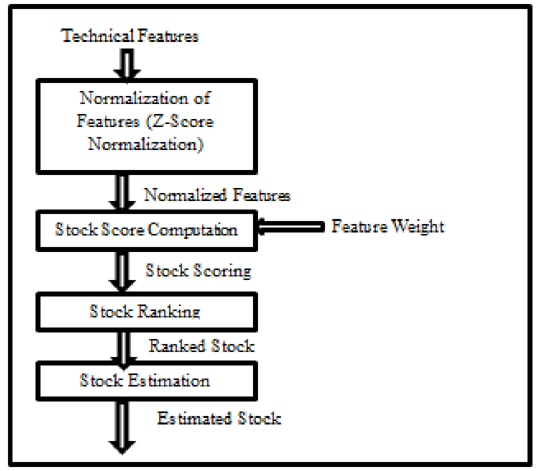

Fig 4: Stock Scoring and Ranking

Technical features are normalized using z-score normalization. Optimized features weight from the differential evolution algorithm. Both the features weight and normalized score are linearly combined to compute the stock score. Based on the score stocks are ranked. The ranked stocks are estimated to predict the next year of the stock.

III.IMPLEMENTATION AND RESULTS

A. Dataset Description

The Nifty Stock Exchange dataset is used for an experiment. The reason for using real-time datasets is to determine the real prediction of the data.

TABLE I
ATTRIBUTES OF NSE DATASET

Attributes	**Description**
Date	Date of stock analyse
Open	Open price on particular day
High	High price on particular day
Low	Low price on particular day
Last	Close price on particular day
Volume	Total stock on particular day

The attributes/characteristics of the stock used in the dataset. The important attributes used for feature constructions are date, open, high, low, close and volume. Open, high, low and close are the price of the stock on a particular day and volume is the amount of the stock on a particular day.

B. *Feature Construction*

ATR	RSI	FI	WILLIAM	SMA	SMADIFF	ROC	ROA
203	99.973	0.91312	80.176	3633.6	1.0561	3633.6	46.842
136.93	99.973	0.01599	79.357	3662.1	1.0195	1.84E+05	15.92
67.783	99.973	0.011558	78.746	3685.8	1.0167	1.87E+05	11.802
61.813	99.973	0.005949	78.477	3702.4	1.0168	1.88E+05	5.3305
70.57	99.974	0.033429	77.163	3730.6	1.0277	1.92E+05	33.101
92.883	99.974	0.001325	77.096	3750.2	1.0102	1.92E+05	1.14
37.986	99.974	0.003752	76.888	3766.3	1.0082	1.93E+05	3.7006
33.975	99.974	0.010153	77.359	3774.2	1.0133	1.91E+05	9.9056
55.656	99.974	0.018845	76.103	3790.1	1.0251	1.96E+05	18.737
96.12	99.975	0.023024	74.725	3812.4	1.0246	2.01E+05	22.269
92.432	99.975	0.006685	75.132	3828.1	1.0135	1.99E+05	6.9323
54.063	99.975	0.007496	74.75	3843.4	1.0144	2.00E+05	6.1878
59.05	99.975	0.003855	74.95	3855.3	1.0197	2.00E+05	0.2349
78.857	99.975	0.020118	73.717	3871.6	1.0237	2.04E+05	19.509

Fig 5: Technical Features

The snapshot of the technical features constructed from input dataset. The above output contains the 8 technical features. The technical features are Average True Range (ATR), Relative Strength Index (RSI), Rate of Change (ROC), Return on Assets (ROA), Simple Moving Average (SMA), Simple Moving Average Difference (SMADIFF), Force Index (FI) and Williams %R. The technical features help to calculate the stock score and to find the right time to buy a stock and sells the stocks.

C. *Features Weight Calculation*

	A	B	C	D	E	F	G	H
1	-0.0011	93.498	0.000325	92.339	20.38	1.0191	415.92	0.735

Fig 6: Feature Weight

The weight is generating by optimizing the technical features using differential evolution algorithm.

D. *Stock Estimation*

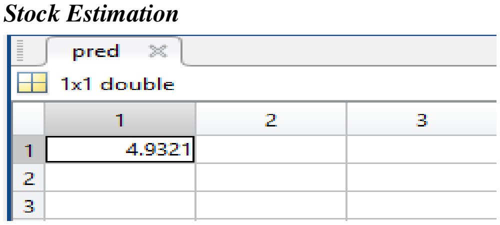

Fig 7: Stock Estimation

Stock estimation calculates mean difference of ranked stock. The difference is the average returns of the next year of the stock.

IV. Conclusions

A system has been constructed for stock selection using Differential Evolution Algorithm. The proposed system is constructed in three steps. In step one features are constructed. In step two constructed features have been utilized for computation of stock scoring. In the third step, stocks are ranked based on the stock score and it has been used for the prediction of the next year stock. Using the Nifty Stock Exchange dataset of India as the study sample, the result shows that the stock selection model is powerful and efficient.

References

[1] C. Cheadle, M.P. Vawter, W.J. Freed and K.G. Becker, "Analysis of microarray data using z score transformation", Journal of Molecular Diagnostics, Vol. 5, No. 2, pp. 73–81, May.2003.

[2] C.F. Huang, B.R. Chang, D.W. Cheng and C.H. Chang (2012), "Feature selection and parameter optimization of a fuzzy-based stock selection model using genetic algorithms", International Journal Fuzzy System, Vol. 14, No. 1, pp. 65–75,2012.

[3] C.F. Huang, T.N. Hsieh, B.R. Chang and C.H. Chang, "A comparative study of stock scoring using regression and genetic-based linear models", IEEE International Conference Granular Computation, pp. 268–273, Nov.2011 .

[4] Lean Yu, Lunchao Hu and Ling Tang, "Stock selection with a novel sigmoid based mixed discrete continuous differential evolution algorithm", IEEE Transaction on Knowledge and Data Engineering, Vol. 28, No.7, pp.1041-4347, 2016.

[5] K. Price, R.M. Storn, and J.A. Lampinen, "Differential Evolution: A Practical Approach to Global Optimization", Springer, Vol. 23, No. 2, pp.37-134, 2006.

[6] J. Wang, M. Wang, P. Li, L. Liu, Z. Zhao, X. Hu, and X. Wu, "Online feature selection with group structure analysis", IEEE Transaction on Knowledge Data Engineering, Vol. 27, No. 11, pp. 3029–3041, Nov.2015.