Implementation and Analysis of Fragment based Mining by using Jstock framework.

Rajesh V. Argiddi,Sulabha S.Apte,Suhas S.Patil

Computer Science Department,
Walchand Institute of Technology, Solapur,India

Abstract— **In this research we propose an idea of using a JStock framework for getting accuracy of stock market prediction algorithms.**
Stock Market is the most unpredictable, popular and a big resource of income all over the world. Many researches are done for predicting the future market prices. Various algorithms lead to varying amount of accuracy. The proposed work will compare the various predictions made by the indicators and algorithms and will lead to a proper testing tool for checking accuracy of the stock market predictors.
We have used JStock an open source software which analyses stock market by using technical indicators. By using the data provided by yahoo finance and the data extraction methods of JStock, we try to analyze the various stock market prediction methods and try to compare their accuracy. The performance of this testing is checked by the dataset provided by yahoo finance of Indian Stock markets and particular Indian Companies.
However, they are less accurate in case of noisy trading, which makes them less appropriate in environments with high uncertainty about true prices for shares.
Keywords— **FITI; Fragment Based Mining, Association mining; Stock Data, Application Software, Algorithm analysis**

I. INTRODUCTION

1.1 What Is Stock market prediction?

A stock market or equity market is a public entity (a loose network of economic transactions, not a physical facility or discrete entity) for the trading of company stock (shares) and derivatives at an agreed price; these are securities listed on a stock exchange as well as those only traded privately.[1]
Stock Market prediction is future prediction of stock market by using historical data and analysing them to achieve patterns for getting probable market prices.
There are various ways of predicting stock prices –
1. Using Technical Indicators.
2. Using historical data.
3. Using stock and related financial rules
4. Using association rules for prediction
5. Using genetic algorithms to get future prices
Such various ways are used to get the probable prices of the required prices of stock market.

II. RELATED WORK

2.1 JStock Open Source Software

JStock software is a GNU Licenced software which uses technical indicators for stock analysis.It extracts data from yahoo finance. JStock is a free stock market software for 26 countries. It provides Stock watchlist, Intraday stock price snapshot, Stock indicator editor,

Stock indicator scanner and Portfolio management. Free SMS/email alert supported.
JStock has a framework that collects data from Yahoo finance API.
It is legal way of collecting data."To retrieve stock information, it is using method provided and allowed by Yahoo.[13]

2.2 Applications

a. Stock Prediction Algorithms
1. Stock Investor gets a fair idea of investment and stock prices.
2. Stock Investors can choose a safer stock with safe price rises when required.
b. Stock Prediction Algorithms Analysis
1. Better view of algorithms accuracy.
2. Proper testing tool for testing algorithms.
3. Can be used for stock analysis also.

III. METHODOLOGY

Proposed idea is to study JStock and create a framework to extract the data from yahoo finance.
The framework will be used to implement fragment based mining algorithm.

3.1: Fragment based mining

Fragment based mining is similar to Granular mining [17].Data sets are fragmented into fragments and hence the input size for algorithm is reduced. The condition and decision attributes are both used to obtain generalized association rules [18].

3.2 Technical Indicators [10] –

Market Indicators are mathematical calculations that are used to predict market values.
 Any class of metrics whose value is derived from generic price activity in a stock or asset. Technical indicators look to predict the future price levels, or simply the general price direction, of a security by looking at past patterns. Examples of common technical indicators include Relative Strength Index, Money Flow Index, Stochastic, MACD and Bollinger Bands.[2]
The indicators are widely used in stock market prediction [3].Lot of research is done in stock prediction and analysis.[4-9].
The indicators are basic techniques used in JStock for scanning and analyzing the various stock simulations.

3.2.1 Moving Average

The Moving Average (MA) is one of the most widely used tools in the technical community. A moving average is a lagging indicator. The importance of the moving average lies in that it tells us the direction of the trend in a smoother

way, it reduces some of the short-term noise in prices, in addition for giving us a clear trading signals.

3.2.2 MACD

The Moving Average Convergence Divergence (MACD) was invented by Gerald Appel, one of the technical analysis veterans. The MACD is consisted of two lines. The first line is the main line and is constructed by subtracting a 26-days exponential moving average from the 12-days exponential moving average. The second line, called signal line is a 9-days exponential moving average of the first line.

3.2.3 Stochastic Oscillator

The Stochastic Oscillator was invented by George Lane, and is of the widely used indicators in technical analysis. It's based on the assumption that whenever an upward move is underway, prices usually close near the highs. When a decline take places, prices usually close near the lows. It's composed of two lines; the fast line is called %K, while the slower one is the %D, which is considered more important, especially during divergences.

%K = C- Ll/ Hh - Ll * 100 (1)

Where C: Close of today, LI: is the lowest low and Hh: is the highest high.

%D = a three days moving average of %K (2)

3.2.4 Relative Strength Indices:

The Relative Strength Index (RSI) was invented by Welles Wilder. It resembles the Stochastic Oscillator despite the different calculations. Wilder created the RSI to show market extremes. It is a bounded oscillator that moves between 0 and 100 and shows overbought and oversold conditions. The RSI calculates the ratio of up closes to down closes over a specific amount of time, as shown in the following formula:

RSI = 100 - (100 / I+RS) (3)

Where RS: is the average of the up closes of the last n-days divided by the average of down closes of the same period.

3.2.5 Average Directional Index

J. Welles Wilder Jr. developed the Average Directional Index (ADX) in order to evaluate the strength of the current trend, be it up or down. It's important to determine whether the market is trending or trading (moving sideways), because certain indicators give more useful results depending on the market doing one or the other. ADX is an oscillator that fluctuates between 0 and 100. Even though the scale is from 0 to 100, readings above 60 are relatively rare. Low readings, below 20, indicate a weak trend and high readings, above 40, indicate a strong trend. The indicator does not grade the trend as bullish or bearish, but merely accesses the strength of the current trend. A reading above 40 can indicate a strong downtrend as well as a strong uptrend.

3.2.6 Outside Price Channel (Number of Days Break Out System)

Next to moving averages, price channels are also extensively used technical trading methods. The price channel is sometimes referred to as "Trading range breakout" or "Support and resistance", and also known as "N-Days Break Out Rule". The fundamental characteristic underlying price channel systems is that market movement to a new high or low suggests a continued trend in the direction established. Thus, all price channels generate trading signals based on a comparison between today's price levels with price levels of some specified number of days in the past. The Outside Price Channel system is analogous to a trading system introduced by Donchian who used only two preceding calendar week's ranges as a channel length. More specifically, this system generates a buy signal anytime the closing price is outside (greater than) the highest price in a channel length (specified time interval), and generates a sell signal anytime the closing price breaks outside (lower than) the lowest price in the price channel.

3.2.7 Use of genetic algorithms for prediction [10]

Genetic algorithms use the basis of evolution. Darwin's theory of survival fittest is basic idea of genetic algorithms (GA). GA is taking the fittest individual from multiple recombination's of the various genes.

Recombination's are biological crossovers where child takes special characteristics of the parents which help them to survive in nature.

Working of GA for stock market techniques-
How Genetic Algorithms Work?

Genetic algorithms are created mathematically using vectors, which are quantities that have direction and magnitude. Parameters for each trading rule are represented with a one dimensional vector that can be thought of as a chromosome in genetic terms. Meanwhile, the values used in each parameter can be thought of as genes, which are then modified using natural selection. For example, a trading rule may involve the use of parameters like Moving Average Convergence-Divergence (MACD), Exponential Moving Average (EMA) and Stochastic. A genetic algorithm would then input values into these parameters with the goal of maximizing net profit. Over time, small changes are introduced and those that make a desirably impact are retained for the next generation.

There are three types of genetic operations that can then be performed:

Crossovers represent the reproduction and biological crossover seen in biology, whereby a child takes on certain characteristics of its parents.

Mutations represent biological mutation and are used to maintain genetic diversity from one generation of a population to the next by introducing random small changes. Selections are the stage at which individual genomes are chosen from a population for later breeding (recombination or crossover).

These three operators are then used in a five-step process:

1. Initialize a random population, where each chromosome is n-length, with n being the number of parameters. That is, a random number of parameters are established with n elements each.
2. Select the chromosomes, or parameters, that increase desirable results (presumably net profit).
3. Apply mutation or crossover operators to the selected parents and generate an offspring.
4. Recombine the offspring and the current population to form a new population with the selection operator.
5. Repeat steps two to four.

Over time, this process will result in increasingly favorable chromosomes (or, parameters) for use in a trading rule. The process is then terminated when a stopping criteria is met, which can include running time, fitness, number of generations or other criteria.[11]

IV. FRAGMENT BASED MINING

Fragment based mining is based on two tier fragmentation, condition fragments and decision fragments. This approach tries to solve the problem of FITI (First intra then Inter). FITI is using algorithm based on the item joins [15].

Hence the problem is solved by the fragment based approach that uses correct width of the sliding window: for example, it may cover only decision attributes, where the width and the length of sliding refer to number of attributes and the intervals of the transactions, respectively. The fragment-based approach is different from FITI algorithm: it uses the sliding windows to separate attributes into layers, and uses small sliding windows for certain layers, and uses small sliding windows for certain layers, based on users' constraints. For example consider there are eight attributes. The sliding window of FITI has all its eight attributes but fragment based approach has further divided it into two layers, condition layers (C1, C2, C3 and C4) and decision layers (D1, D2, D3 and D4)[16]. The sliding window is set only for the decision attributes, as shown in the fig.1.1 and fig 1.2.

ID	C1	C2	C3	C4	D1	D2	D3	D4
1								
2			Window 1					
3								
4								
5								
6								
7								
8								
9								

Fig.1.1 FITI

ID	C1	C2	C3	C4	D1	D2	D3	D4
1								
2					Window 1			
3								
4								
5								
6								
7								
8								
9								

Fig.2 Fragment Based Approach

V. PROPOSED ARCHITECTURE

The algorithm approach is consisting of two stages. First we will study the JStock software. Then we will build a JStock framework which will extract the Stock Data. We will then apply Fragmentation mining algorithm. Then we will analyze the results. Flow diagram shown in 2.1.
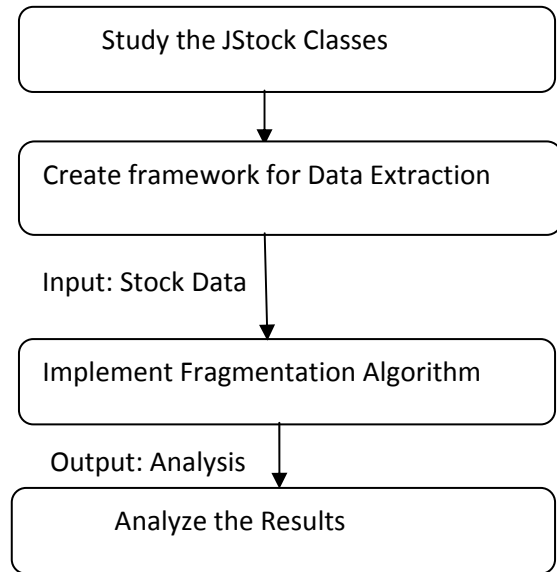
Fig.2.1 Implementation flow.

We will follow following steps to implement our algorithm-

1.In this method we will perform following calculation, we will represent last 20 days record in graph by using granule method.

ID	Date	B1	B2	B3	SUM	99.7% * SUM	100. * SI
1	3/01/2002	18.88	7.84	2.9	29.62	29.53	29.?
2	4/01/2002	18.8	7.75	2.86	29.41	29.32	29.4
3	7/01/2002	19.03	7.75	2.86	29.64	29.55	29.?
4	8/01/2002	19.11	7.78	2.84	29.73	29.64	29.8
5	9/01/2002	18.88	7.68	2.76	29.32	29.23	29.4
6	10/01/2002	18.97	7.61	2.72	29.3	29.21	29.?
7	11/01/2002	19.0	7.82	2.8	29.62	29.53	29.?
8	14/01/2002	18.9	7.83	2.8	29.53	29.44	29.6
9	15/01/2002	18.7	7.65	2.81	29.16	29.07	29.?
10	16/01/2002	18.84	7.5	2.84	29.18	29.09	29.?
11	17/01/2002	18.79	7.31	2.87	28.97	28.88	29.0
12	18/01/2002	18.63	7.28	2.82	28.73	28.64	28.8
13	21/01/2002	18.63	7.4	2.78	28.81	28.72	28.8
14	22/01/2002	18.51	7.4	2.78	28.69	28.60	28.?
15	23/01/2002	18.5	7.4	2.8	28.7	28.61	28.?
16	24/01/2002	19.0	7.4	2.94	29.34	29.25	29.4
17	25/01/2002	19.33	7.42	2.99	29.74	29.65	29.8
18	29/01/2002	19.35	7.37	2.9	29.62	29.53	29.?
19	30/01/2002	19.15	7.5	2.86	29.51	29.42	29.?

2. We will decide window length according to the condition

ID	Date	SUM	99.7% * SUM	100.3% * SUM	△ SUM
1	3/01/2002	29.62	29.53	2 9.70	1
2	4/01/2002	29.41		**W1**	-
3	7/01/2002	29.64			-
4	8/01/2002	29.73			-
……		…	…	…	
6	10/01/2002	29.3	29.21		-1
7	11/01/2002	29.62		**W7**	-
8	14/01/2002	29.53			-
9	15/01/2002	29.16			-
……		…	…	…	
11	17/01/2002	28.97	28.88		0
12	18/01/2002	28.73		**W12**	-
13	21/01/2002	28.81			-
14	22/01/2002	28.69			-
……		…	…	…	

Fig Window length.

3. We will decide the status of △(sum) according to range

ID	Date	SUM	99.7% * SUM	100.3% * SUM	△ SUM
1	3/01/2002	29.6	29.53	29.70	1
2	4/01/2002	29.4	29.32	29.49	1
3	7/01/2002	29.6	29.55	29.72	1
4	8/01/2002	29.7	29.64	29.81	-1
5	9/01/2002	29.3	29.23	29.40	1
6	10/1/2002	29.3	29.21	29.38	1
7	11/1/2002	29.62	29.53	29.70	-1
8	14/1/2002	29.53	29.44	29.61	1
9	15/1/2002	29.1	29.07	29.24	-1
10	16/1/2002	29.1	29.09	29.26	-1
11	17/1/2002	28.9	28.88	29.05	-1
12	18/1/2002	28.7	28.64	28.81	
13	21/1/2002	28.8	28.72	28.89	
14	22/1/2002	28.6	28.60	28.77	
15	23/1/2002	28.7	28.61	28.78	

4. We will create graph for data representation accordingly.
5. We can analyze the results finally.

CONCLUSIONS

With the help of this approach we propose to use the JStock framework to extract dynamic Stock data and use the framework as a testing tool for algorithms and also new technical indicators.

REFERENCES

[1] Wikipedia page definition for stock market
http://en.wikipedia.org/wiki/Stock_market
[2] Investopedia definition
http://www.investopedia.com/terms/t/technicalindicator.asp
[3] Shipra Banik, A.F.M. Khodadad Khan and Mohammad Anwer Dhaka tock Market Timing Decisions by Hybrid Machine Learning Technique IEEE 2012
[4] J. Yao and J.P. Herbert, "Financial time-series analysis with rough sets", Applied Soft Computing, pp.1000-1007, 2009.
[5] L. Shen and H.T. Loh, "Applying rough sets to market timing decisions", Decision Support Systems, pp.583-597, 2004.
[6] S.M. Shamsuddin, S.H. Jaaman and M. Darus, "Neuro-rough trading rules for mining Kuala Lumpur composite index", European Journal of Scientific Research, pp.278-286, 2009.
[7] S.H. Jaaman, S.M. Shamsuddin, B. Yusob and M. Ismail, "A predictive model construction applying rough set methodology for Malaysian stock market returns", International Research Journal of Finance and Economics, pp.211-218, 2009.
[8] C.F. Tsai and S.P. Wang,"Stock price forecasting by hybrid machine learning techniques", Proceedings of the International MultiConference of Engineers and Computer Scientists, Hongkong, 2009.
[9] F.E.H Tay and L. Shen, "Economic and financial prediction using rough sets model", European Journal of Operational Research, pp.643-661, 2002.
[10] Fahima A. Badawy GENETIC ALGORITHMS FOR PREDICTING THE EGYPTIAN STOCK MARKET
[11]Investopedia Genetic algorithms http://www.investopedia.com/articles/financial-theory/11/using-genetic-algorithms-forecast-financial-markets.asp
[12] JStock Web Links home page http://www.jstock.org/ http://www.ohloh.net/p/jstock
[13] JStock Legal FAQ http://www.jstock.org/help_faq.html#legal
[14]http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/Clustering_Parameters/Euclidean_and_Euclidean_Squared_Distance_Metrics.htm
[15] A.K.H. Tung, H. L., J. Han, and L. Feng (2003). "Efficient mining of intertransaction association rules." IEEE Transactions on Knowledge and Data Engineering 15(1): 43-56
[16] Y. Li and N. Zhong (2003). Interpretations of Association Rules by Granular Computing. The Third IEEE International Conference on Data Mining.
[17] Wanzhong Yang, "Granule Based Knowledge Representation for Intra Inter Transaction Association Mining", Queensland University of Technology, July 2009.
[18] Fragment Based Approach to Forecast Association Rules from Indian IT Stock Transaction Data IJCSIT Volume 3 (2), '2012'