ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

Genetic based stock market Prediction

Prof R.V.Patil, Ganesh Yuvraj Tamhane, Jitendra S Dhage, Ashish B Gawande, Suraj U Mogal.

Abstract— this paper proposes a system that will provide predictions about the share market, which will follow two main phases, which are fragment based association mining and further on more optimization and prediction will be provided by genetic algorithm. The major advantage of using fragment based mining is that, it groups all the attributes once and performs operations group wise instead of single attributes which results in more generalized rules which are furtherly highly optimized using genetic algorithm as its time space complexity is less than any other algorithm and provide prediction of small scale companies based on transaction data of large scale as well as small scale companies.

I. INTRODUCTION

The prediction of the stock market is a fascinating subject. It changes the lives of investors on a daily basis based on the decisions they make whether to buy or sell shares. The stock market is most important resource for companies to raise money. This allows businesses to be publicly traded by selling shares of ownership of the company in a public market. Research shows that the price of shares and other assets is an important part of the dynamics of economic activity and can influence social mood. In fact, the stock market is often considered the primary signal for country's economic strength and development. Leading share prices tends to be associated with increased business investment and vice versa. Therefore, central banks tend to keep an eye on the control and behavior of the stock market. Exchanges also act as the clearing house for each transaction, meaning that they collect and deliver the shares, and guarantee payment to the seller of a security. This eliminates the risk to an individual buyer or seller that the counter party could default on the transaction. The smooth functioning of all these activities facilitates economic growth, lower costs; promote the production of goods and services as well as employment. In this way the financial system contributes to increased prosperity.

Stock market Price prediction using data mining is one of the most fascinating issues to be investigated and it is one of the important issues of stock market research over the past decade. However, determining the best time to buy, sell or hold a stock remains very difficult because there are lots of factors that may influence the stock market prices like fundamental, Technical indexes, unknown factors.

A. Overview

Association rule mining is the process of discovering knowledge about relationship between items or events [1, 2, 3], which ideally can be applied to estimate upcoming states of a dynamic system, e.g. movement of stock prices in a stock exchange market. The word "movement" in this study means the direction of a stock price, i.e. up, down, or stable (unchanged). An example of a rule that can be generated from traditional association rule mining is given below:

R1: When the prices of IBM and SUN go up, at 80 percent of probability the price of Microsoft goes up on the same day. In the given example extracted rule might be of help for investors to discover relationship between movements of company stock prices, however its role in stock price prediction is limited. Conversely, market participants, in particular traders, demanded a more detail form of expression or rule as follows:

R2: When the prices of IBM and SUN go up, Microsoft's will most likely (80 percent of probability) go up the next day. With such kind of a rule, it would be possible for investors or stock traders to predict the stock price movements on the next day.

The process of extracting such rules from a set of transaction or inter-transaction data set is known as the association rule-mining process. One basic technique used in the association rule-mining algorithm is the apriori algorithm. The principle of apriori algorithm is to reduce the number of candidates with pruning techniques, reduce the number of comparisons, so it is not needed to match every candidate against every transaction and anti-monotone [4, 5]. The main purpose of this study is to forecast and analyze company stock price movements in the Indonesia Stock Exchange market (IDX) by implementing the association rule-mining algorithm to extract rules of relationship between movements of company stock prices from time to time. In addition, the study also examines and evaluates several factors that affect the number of rules generated. Throughout conducted

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

experiment, it is expected to have more number of rules generated as by being able to accomplish such task more conditions between movements of company stock process can be modelled better.

B. Brief Description

The prediction of the stock market is a fascinating subject. It changes the lives of investors on a daily basis based on the decisions they make whether to buy or sell shares. The stock market is most important resource for companies to raise money. This allows businesses to be publicly traded by selling shares of ownership of the company in a public market. Examination demonstrates that the cost of shares and different resources is a critical piece of the progress of financial action and can impact social mind-set. Indeed, the stock exchange is regularly viewed as the essential sign for nation's monetary quality and improvement. Driving offer costs has a tendency to be connected with expanded business venture and the other way around. Along these lines, national banks have a tendency to watch the control and conduct of the stock exchange. Trades likewise go about as a clearing house for every exchange importance that they gather and convey the shares and guarantee installment to the dealer of security. This wipes out the danger to an individual purchaser or merchant that the counter party could default on the exchange. The smooth working of all these exercises encourages financial development, lower expenses; advance the creation of merchandise and administrations and occupation. Thus the monetary framework adds to expanded prosperity. Stock exchange Price expectation utilizing information mining is a stand out amongst the most interesting issues to be researched and it is one of the essential issues of securities exchange inquire about over the previous decade. In any case, deciding the best time to purchase, offer or hold a stock stays exceptionally troublesome on the grounds that there are loads of components that may impact stocks costs like basic, Technical files, obscure factors. Information mining system like Association tenet mining (ARM) concentrates on discovering most successive thing sets and relating affiliation rules. Section based principle mining system by taking simply chronicled datasets as data is proposed. In any case, this calculation produces a greatly vast number of affiliation standards, regularly in hundreds or even thousands. Further, the affiliation standards are once in a while huge. It is about inconceivable for the end clients to see effortlessly. Subsequently, the clients need to utilize an attempt and -lapse way to get suitable number of standard.

C. Purpose

The purpose of this document is to specify requirements for software named as Genetic based stock Market prediction is to provide prediction of stock market. Other objectives often include: maintainability, security, and consistent end-user documentation, adequate comments in code, complete unit tests, and scalability.

D. Steps

As stated in previous points stock prediction in this proposed system can be performed using (Fragment based mining) and further on applying genetic algorithm to provide the rules with maximum fitness. Different methods which are already available and used for the purpose of the finding of prediction rules causes failure in the prediction when it comes to real time implementation or as the data increases Therefore this paper proposes a system which will be using fragment based mining and genetic algorithm to provide more appropriate prediction. This technique is divided in following steps and works as follows- This technique is divided in following steps and works as follows-

1 Data collection-

The very first thing which is required is up-to date data rather say up-to time data of large scale companies and small companies should be maintained in the tabular form with its date, from online finance sites like yahoo finance.

2 Data Extraction

Second thing when prediction is require of some particular companies then data of that companies should be extracted from the collected data for the further processing. The extracted tabular data will contain transaction data of large scale companies and transaction data of small scale companies.

3. Datpreprocessing

In share market the difference between the two transactions represents the status level of attributes. Let Δ be the difference from attribute values among inter transactions Assume 1, 0,-1 illustrate increase, neutral and decrease respectively .Each transaction represents the change in previous date transaction.

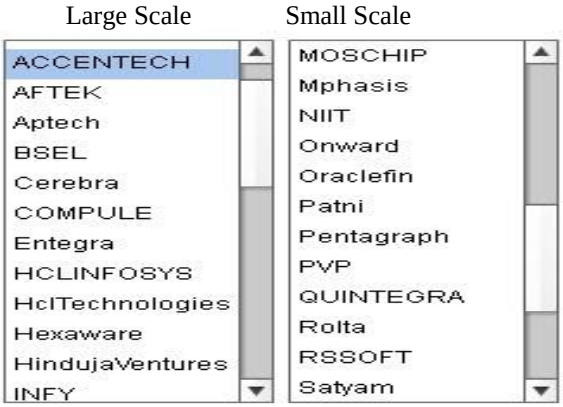
ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

4. ***Fragment based mining***

Fragment based approach divides the attributes in two tiers:
-Small scale and Large scale SUM functions.

After this we have to find differences for attribute values among inter-transaction, Assume 1, 0 illustrate increase and decrease respectively. Now according to our approach we will consider only those transactions whose both small scale and large scale SUM is same i.e. both are 1, 1 or 0, 0 respectively. This we do because we are only interested in finding the association if both small and large scale companies increase or decrease at the same time. So the original transaction table will get minimized. This showed that fragment based mining algorithm gets accurate results with less time and space complexity as compared to FITI algorithm.

As in granule mining, fragment based approach fragments the data sets into fragments for processing thereby reducing the input size of data sets fed to the algorithm. In contrast to granule mining, in fragment based mining the condition and decision attributes are summed for obtaining generalized association rules.



The data of selected companies is extracted as shown in following figure-

ID	*Date*	*A1*	*A2*	*A3*	*A4*	*B1*	*B2*	*B3*
0	2/01/2002	15.7	17.78	31.75	29.9	18.45	7.7	2.82
1	3/01/2002	15.84	17.96	31.89	30.09	18.88	7.84	2.9
2	4/01/2002	15.71	18.0	32.08	29.92	18.8	7.75	2.86
3	7/01/2002	15.5	17.85	31.88	29.9	19.03	7.75	2.86
4	8/01/2002	15.35	17.69	31.8	29.95	19.11	7.78	2.84
5	9/01/2002	15.15	17.25	31.34	29.98	18.88	7.68	2.76
6	10/01/2002	14.97	17.2	31.0	29.50	18.97	7.61	2.72
7	11/01/2002	15.15	17.35	31.1	29.72	19.0	7.82	2.8

Now pre-processing is done on the extracted data by using formula Δ=next transaction –previous transaction.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

ID	A1	A2	A3	A4	B1	B2	B3
1	1	1	1	1	1	1	1
2	-1	1	1	-1	-1	-1	-1
3	-1	-1	-1	-1	1	0	0
4	-1	-1	-1	1	1	1	-1
5	-1	-1	-1	1	-1	-1	-1
6	-1	-1	-1	-1	1	-1	-1
7	1	1	1	1	1	1	1

To convert decision attributes into decision granules, aggregation SUM function is calculated and ΔSUM is Calculated using Sliding window function.

ID	Date	SUM	99.7% * SUM	100.3% * SUM	ΔSUM
1	3/01/2002	29.62	29.53	29.70	1
2	4/01/2002	29.41	29.32	29.49	1
3	7/01/2002	29.64	29.55	29.72	1
4	8/01/2002	29.73	29.64	29.81	-1
5	9/01/2002	29.32	29.23	29.40	1
6	10/01/2002	29.3	29.21	29.38	1
7	11/01/2002	29.62	29.53	29.70	-1

By combining both the tables of large scale and small scale companies we form the table based on the covering set-

ID	A1	A2	A3	A4	ΔSUM =1	ΔSUM =0	ΔSUM = -1	N	Covering Set
1	$a_{1,1}$	$a_{2,1}$	$a_{3,1}$	$a_{4,1}$	1		1	2	(1, 7)
2	$a_{1,1}$	$a_{2,1}$	$a_{3,1}$	$a_{4,3}$			1	1	(11)
3	$a_{1,1}$	$a_{2,3}$	$a_{3,3}$	$a_{4,1}$	1		1	2	(9, 15)
4	$a_{1,1}$	$a_{2,3}$	$a_{3,1}$	$a_{4,3}$	1			1	(16)
5	$a_{1,2}$	$a_{2,3}$	$a_{3,1}$	$a_{4,1}$	1			1	(8)
6	$a_{1,3}$	$a_{2,1}$	$a_{3,1}$	$a_{4,3}$	2	1	1	4	(2,10,12,14)
7	$a_{1,3}$	$a_{2,3}$	$a_{3,1}$	$a_{4,1}$	1			1	(13)
8	$a_{1,3}$	$a_{2,3}$	$a_{3,3}$	$a_{4,1}$	1		1	2	(4,5)
9	$a_{1,3}$	$a_{2,3}$	$a_{3,3}$	$a_{4,3}$	2			2	(3, 6)

Finally after completing all above procedure, rules will be generated, and their confidence is calculated Cerebra (↓), BSEL (↑), Onward (↑), Jetking (↓) => Satyam, HCL, Patni (↑)

……. Association Rule (1)

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

Confidence: 89.10
Cerebra (↓), BSEL (↓), Onward (↓), Jetking (↑) => Satyam, HCL, Patni (↑)

……. Association Rule (2)

Confidence: 85.98
Cerebra (↑), BSEL (↓), Onward (↓), Jetking (↓) => Satyam, HCL, Patni (↑)

……. Association Rule (3)

Confidence: 84.55
Cerebra (↓), BSEL (↓), Onward (↓), Jetking (↓) => Satyam, HCL, Patni (↑)

……. Association Rule (4)

Confidence: 83.75

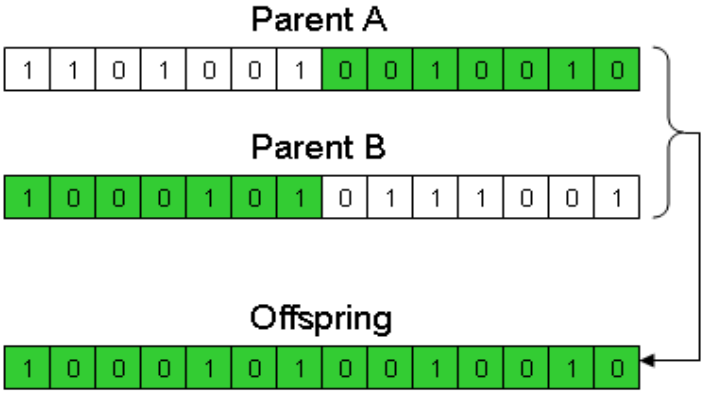
Further optimization is performed using genetic algorithm by selecting rules having maximum confidence.

Genetic algorithm for optimization

Genetic algorithm (GA) is a search heuristic that mimics the process of natural Evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms, which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. The genetic algorithms are important when discovering association rules because they work with global search to discover the set of items frequency and they are less complex than other algorithms often used in data mining. The genetic algorithms for discovery of association rules have been put into practice in real problems.

Genetic Operation: The Genetic operators determine the search capability and convergence of the algorithm. Genetic operators hold the selection crossover and mutation on the population and generate the new population.

1Select operation: In this algorithm it restores each chromosome in the population
To the corresponding rule, and then calculate selection probability pi for each rule.
2Crossover operation: In which multi point crossover are used. It classifies the domain of each attribute into a group and classifies the cut point of each continuous attributes into one group and the crossover carried out between the corresponding groups of two individuals by a certain rate. Crossover is performed by selecting a random gene along the length of the chromosomes and swapping all the genes after that point. For example, given two chromosomes, choose a random bit along the length, say at position 9, and swap all the bits after that point.

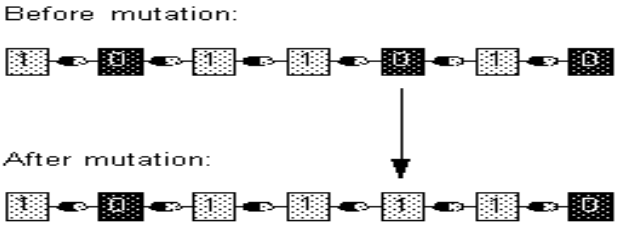


3Mutation operation: Mutation alters the new solutions so as to add stochasticity

In the search for better solutions. This is the chance that a bit within a chromosome

Will be swapped (0 becomes 1, 1 becomes 0).[5] Whenever chromosomes are chosen from the population the algorithm first checks to see if crossover should be applied and then the algorithm iterates down the length of each chromosome mutating the bits if applicable.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016



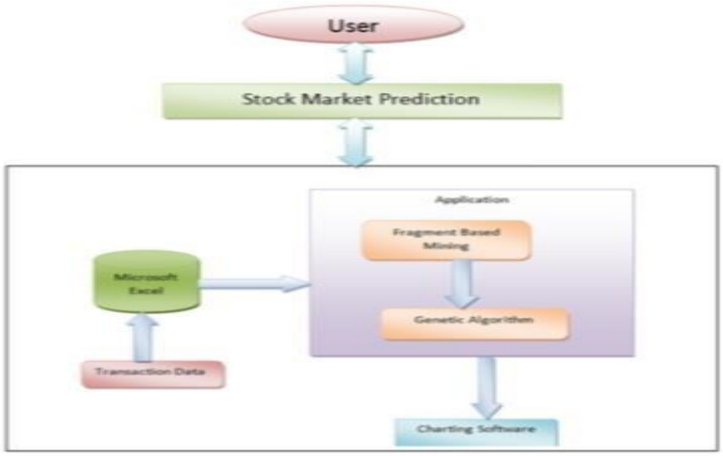
Apriori Algorithm Developed by Agarwal and Srikant 1994 Innovative way to find association rules, Apriori is designed to operate on databases containing transactions. The algorithm attempts to find subsets which are common to at least a minimum number confidence threshold of the item-sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time a step known as candidate generation, and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a hash tree structure to count candidate item sets efficiently. But it is costly to handle a huge number of candidate sets .Also it is tedious to repeatedly scan the database and check a large set of candidates by pattern matching.

FITI (First Intra then Inter) The FITI algorithm is based on the following property, a large inter-transaction item-set must be made up of large intra-transaction item-sets, which means that for an item-set to be large in inter-transaction association rule mining, it also has to be large using traditional intra-transaction rule mining methods. By using this property, the complexity of the mining process can be reduced, and mining inter transaction association rules can be performed in a reasonable amount of time. First FITI introduces a parameter called max span (i.e. sliding window size) denoted as w. This parameter is used in the mining of association rules, and only rules spanning less than or equal to w transactions will be mined. Second, every sliding window in the database forms a mega transaction. A mega transaction in a sliding window W is defined as the set of items W, appended with the sub window number of each item. The items in the mega transactions are called extended items. Txy is the set of mega transactions that contain the set of extended items X, Y, and Tx is the set of mega transactions that contain X. The support of an inter-transaction association rule X=> Y is then defined as- Support = |Txy| /S,

Confidence = |Txy|/|Tx|.

The weaknesses in the FITI approaches such as time and space involved in processing the data is more. In FITI approach it is difficult to process an information table with many attributes and long intervals for inter transaction associations. This results into large amount of time and cost in processing the data. Fragment based mining groups all the attributes once and performs the operation group wise instead of single attribute, which results into more generalized rules.

II. SYSTEM DESIGN



ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 5, Issue 2, March 2016

REFERENCES

[1] X. Wu, X. Zhu, G. Q. Wu, et al., "Data mining with big data," IEEE Trans. on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, January 2014.

[2] Z. Zheng, J. Zhu, M. R. Lyu. "Service-generated Big Data and Big Data-as-a-Service: An Overview," in Proc. IEEE BigData, pp. 403-410, October 2013.

[3] A. Bellogín, I. Cantador, F. Díez, et al., "An empirical comparison of social, collaborative filtering, and hybrid recommenders," ACM Trans. on Intelligent Systems and Technology, vol. 4, no. 1, pp. 1-37, January 2013.

[4] R. S. Sandeep, C. Vinay, S. M. Hemant, "Strength and Accuracy Analysis of Affix Removal Stemming Algorithms," International Journal of Computer Science and Information Technologies, vol. 4, no. 2, pp. 265-269, April 2013.

[5] V. Gupta, G. S. Lehal, "A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages," Journal of Emerging Technologies in Web Intelligence, vol. 5, no. 2, pp. 157-161, May 2013.

AUTHOR BIOGRAPHY

Prof. R.V.Patil Lecturer at College of engg Manjari BK. Pune, Maharashtra, India.

Ganesh Y Tamhane He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.

Jitendra Dhage He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.

Suraj Mogal He is pursuing BE (COPMUTER) degree from College of Engg Manjari BK. Pune, Maharashtra, India.