



Genetic Based Share Market Prediction

Harshada Ashok Kore¹ Samata Ajitkumar Gandhi² Gaikwad Neha Pandharinath³
S.B.Patil college of
Engineering, India

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 further 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.

Keywords— Data mining, Fragment based mining, Association rule mining, optimization, Genetic algorithm.

1, 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.[8]

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.[6]

Data mining technique like Association rule mining (ARM) focuses on finding most frequent item sets and corresponding association rules. Fragment based rule mining method by taking just historical datasets as input is proposed. But, this algorithm generates an extremely large number of association rules, often in



hundreds or even thousands. Further, the association rules are sometimes very large. It is nearly impossible for the end users to understand easily. Consequently, the users have to use a try and -error approach to get suitable number of rule.

2. LITERATURE REVIEW-

| Sr No | Name of Algorithm | Behavioural characteristics | Drawbacks |
|-------|--|--|--|
| 1 | Apriori | ➤ Finds subsets common to minimum number confidence threshold of item-sets. | ➤ Costly to handle huge amount of candidate set. |
| 2. | FITI | ➤ Introduces parameter called maxspan (w),used in the mining of association rules, and only rules spanning less than or equal to w transactions will be mined. | ➤ Increased time and space complexity |
| 3. | Artificial Neural Network | ➤ Detects complex non-linear relationship between dependent and independent variables. | ➤ Black box nature Greater computational burden |
| 4. | Fragment based mining Excluding Genetic algorithm | ➤ Divides the attributes into tiers | ➤ Creates large number of generalized rules. |

2.1 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. [10]

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.[2]

2.2 FITI (First Intra then Inter)

The FITI algorithm [11] 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 maxspan (or sliding window size), denoted 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.[3]

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.

T_{xy} is the set of mega transactions that contain the set of extended items X, Y , and T_x is the set of mega transactions that contain X .

The support of an inter-transaction association rule $X \Rightarrow Y$ is then defined as " Support = $|T_{xy}| / S$, Confidence = $|T_{xy}| / |T_x|$.

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.[2]

Fragment based mining groups all the attributes once and performs the operation group wise instead of single attribute, which results into more generalized rules.

3.PROPOSED SYSTEM-

Methodology-

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.[1]

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-

1 .Data collection-

The very first thing which is required is upto date data rather say upto 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. Data preprocessing-

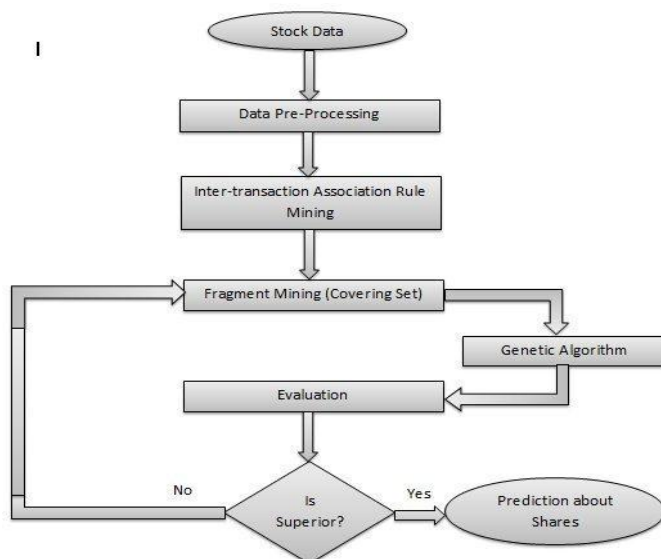


Fig 1.Flow of Project

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

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.[9]

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. After performing fragment based mining the output generalized rules are provided to genetic algorithm.

Consider the following example-

User is asked to select large scale and small scale companies from given list –

Large-Scale

Small-Scal

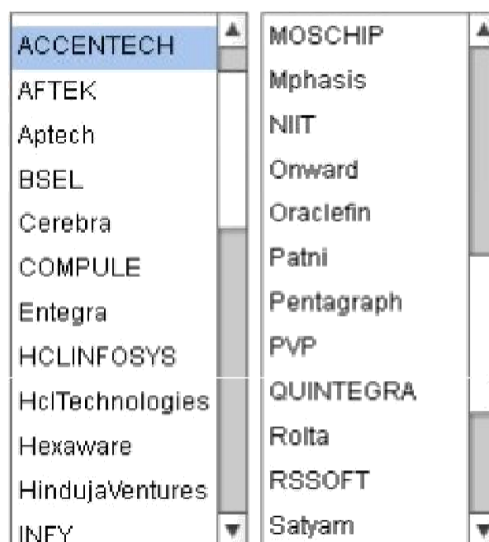


Fig2.selection of companies



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

| <i>ID</i> | <i>Date</i> | <i>A1</i> | <i>A2</i> | <i>A3</i> | <i>A4</i> | <i>B1</i> | <i>B2</i> | <i>B3</i> |
|-----------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 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 |

Fig.3.Extracted data of selected companies

Now pre-processing is done on the extracted data by using formula

$\Delta = \text{next transaction} - \text{previous transaction}$.

| <i>ID</i> | <i>A1</i> | <i>A2</i> | <i>A3</i> | <i>A4</i> | <i>B1</i> | <i>B2</i> | <i>B3</i> |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 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 |

Fig.4.-Preprocessed transactions



| 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 |

Fig-5.Sum function for decision attributes

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) |

Fig.6.Fragment based inter transaction association

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)

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.

5. 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.[4]

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.[5]

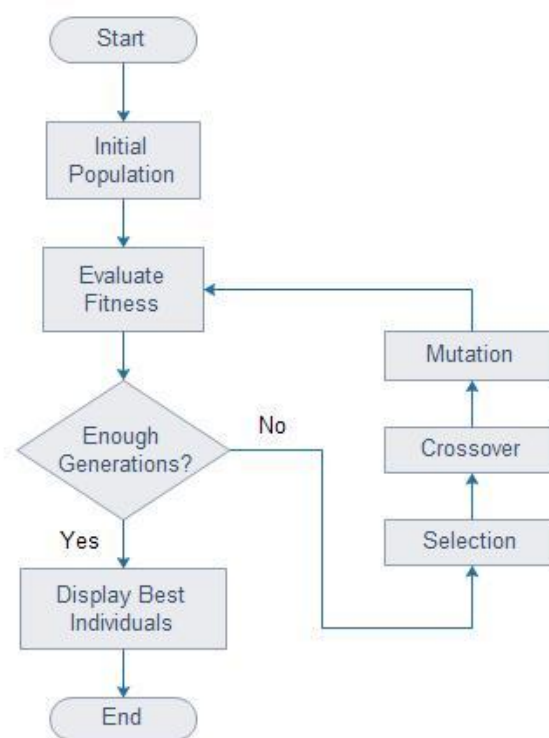


Fig.7.Steps in Genetic algorithm.

5.1Select operation: In this algorithm it restores each chromosome in the population To the corresponding rule, and then calculate selection probability p_i for each rule.

5.2Crossover operation: In which multi point crossover are used. It classifies the domain of each



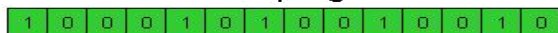
Parent A



Parent B



Offspring



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

Fig.7.-Crossover

5.3 Mutation 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.

Before mutation:



After mutation:



Fig.8.Mutation

4.CONCLUSION AND FUTHER WORK

The proposed method that is idea of using Fragment based mining along with genetic algorithm can help providing prediction of share market with considerably good accuracy. Additionally the extended work required is using more appropriated algorithm for each phase of genetic algorithm by proper analyzation

REFERENCES-

[1]Harya Widiputra ,Bagus Pahlevi “Inter-transaction Association Rule Mining in

the Indonesia Stock Exchange Market”International Conference on Uncertainty Reasoning and Knowledge Engineering,2012.



- [2] Rajesh V. Argiddi, S. S. Apte, "Future Trend Prediction of Indian IT Stock Market using Association Rule Mining of Transaction data", IJCA, Volume 39– No.10, February 2012.
- [3] Rajesh V. Argiddi, Sulabha S. Apte, "Fragment Based Approach to Forecast Association Rules from Indian IT Stock Transaction Data" IJCSIT, Vol. 3 (2) , 2012
- [4] Anubha Sharma, Nirupma Tivari, "A Survey of Association Rule Mining Using Genetic Algorithm" IJCAIT, Vol. I, Issue II, August 2012.
- [5] Anandhavalli M., Suraj Kumar Sudhanshu, "Optimized association rule mining using genetic algorithm" Bioinfo Publications, Advances in Information Mining, ISSN: 0975–3265, Volume 1, Issue 2, 2009. [6] Rajesh V. Argiddi, Sulabha S. Apte, "A Study Of Association Rule Mining In Fragmented Item-Sets For Prediction Of Transactions Outcome In Stock Trading Systems", IJCET, IAEME, Volume 3, Issue 2, July- September (2012).
- [7] Wanzhong Yang, "Granule Based Knowledge Representation for Intra and Inter Transaction Association Mining, Queensland University of Technology", July 2009.
- [8] Rajesh V. Argiddi, Sulabha S. Apte, "AN EVOLUTIONARY FRAGMENT MINING APPROACH TO EXTRACT STOCK MARKET BEHAVIOR FOR INVESTMENT PORTFOLIO" Volume 4, Issue 5, September - October (2013), © IAEME, July 2009.
- [9] J. Hipp, U. Gntzer, and G. Nakhaeizade, "Algorithms for association rule mining a general survey and comparison", ACM SIGKDD Explorations Newsletter, New York, USA, 2000.
- [10] Professor Lee "Apriori Algorithm Review for Finals" Spring 2007.
- [11] Ole Kristian Fivelstad "Temporal Text Mining" Norwegian University of Science and Technology, June 2007.