

A System Design: Correlated Pattern Mining in Intraday Stock Trading Transactions

Mr. R.V. Argiddi¹ Miss. Rajeshwari.K.Gundla²

¹Associate Professor

^{1,2}Department of Computer Science & Engineering

^{1,2}Walchand Institute of Technology, Solapur, India

Abstract— The search for association rules in market is one of the most prominent studied problems in Data Mining. This problem often referred as association rule mining and also roughly as correlation mining. Association rules are intended to identify patterns of the type: “A customer purchasing item X often also purchases item Y.” These associations identified further can be generalized as correlations and dependencies among the item sets. Highly motivated by the inferential nature of association rule mining and its implementation feasible nature for applications beyond market basket data, we develop the notion of correlated patterns in Stock Market Trading and further focus around trading patterns during In traday period. A typical Inter transaction association rule mining in Stock Market is of form “If company ABC’s stocks go up on day one then company XYZ’s stocks goes down on day two”, these inter transaction rules highlight associations between different transactions. We in our work propose a framework which would find and describe relevance of inter-transaction associations mined as intraday patterns to transactions with inter transaction associations mined across inter day patterns. We use modular approach for designing the system for its easy integration and scalability. This proposed modular framework will be put to test by using data obtained from Indian Stock Market using stock feed extractors for Intraday Trading volume data for Stocks.

Key words: Stock Market, Association Rule Mining, Intraday Transactions, Hadoop

I. INTRODUCTION

Financial institutions all over world as like stock markets generate humungous datasets that build a foundation for approaching these enormously complex and dynamic problems of forecasting and analytical insights with data mining tools.

Stock market is a central place for trading stocks. To better pinpoint the motivation for our proposed system the following overview assists the reader in better understanding of introductory terminologies in stock Market.

Analysis of Stock market data involves many attributes, far more than traders can readily and simply understand and interpret them. Traders nonetheless attempt to determine relationships between the data attributes that can lead to profitable trading of financial instruments. Traders apply two types of complementary analysis to trade on the stock market: technical and fundamental analysis.

Fundamental analysis: traders study the underlying factors that determine the price of a financial instrument. For example, factors such as a company’s profit, market sector, or potential growth can influence the share price. Traders consider these factors more crucial than global concerns such as the general economic trend. Traders have traditionally used fundamental analysis to trade the market.

Technical analysis: is “the study of behavior of market participants, as reflected in price, volume, and open interest for a financial market, in order to identify stages in the development of price trends.” In technical analysis, traders ignore the underlying factors that determine price and assume that the price of a financial instrument already quantifies these underlying factors. Technical analysis relies on patterns found directly in the stock data. Because this work relies on the user’s finding patterns directly in the data, it is based on technical analysis. Many traditional analysts don’t support the assumptions made by technical analysts, because it ignores the underlying market factors on which stock prices are based and so is thought to be less reliable [14].

In data mining, association rule also known as correlation mining using support and confidence factors is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro describes analyzing and presenting strong rules discovered in databases using different measures of interestingness [1]. Based on the concept of strong rules, Agrawal et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) systems in supermarkets [2].

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules shows attributed value conditions that occur frequently together in a given dataset. Mining association rules on large datasets has received considerable attention in recent years. Association rules are useful for determining correlations between attributes of a relation and have applications in marketing, financial, and retail sectors. Furthermore, optimized association rules are an effective way to focus on the most interesting characteristics involving certain attributes.

Optimized association rules are permitted to contain uninstantiated attributes and the problem is to determine instantiations such that either the support or confidence of the rule is maximized. For example, data are collected using bar-code scanners in supermarkets. Such ‘market basket’ databases consist of a large number of transaction records. Each record lists all items bought by a customer on a single purchase transaction. Managers could use this data for adjusting store layouts, cross-selling, promotions, and catalog design and to identify customer segments based on buying patterns.

II. BACKGROUND

Stock market is a place where the companies and stockholders get revenue. People are trading in the company and it is regular source of income. There are plenty of sources people get the information to make the investment in stock market such as Books, Internet, news and also from

previous experience. Many times market is in unpredictable conditions so proper investment in stock market is problem for investor. So this approach is suitable for prediction in different sectors of the stock market. Stock market allows companies for publicly trade of the business, or raises the capital with selling the shares of company. It provides companies with access to capital and also for investors with a slice of ownership in the company get the profit based on company's future performance. There are different types of transaction based on trading sessions in the stock market intraday transaction-This is transaction within the day. The term intraday is used to describe the trade on markets during regular business hours, such as opposed stocks and ETFs. This is also called as short term investment.

Interday transaction-This is transaction within week or month. Here the investment is for long term.

III. LITERATURE REVIEW

A system called News CATS was proposed, which took news items as an indicator to predict trends in stock trading scenarios [4]. Several researchers like [12] use twitter mood as stock market trend predictor. [5] Was one of the foremost work carried out for inter transaction association rule mining laying foundations for association rule mining in stock markets in broader sense.

To support multidimensional association rules, Goil et al. [8] presented a scalable parallel system with the techniques of OLAP and data mining to calculate support and confidence. Moreover, Nestorov et al. proposed an approach that can keep all query- processing within the data warehouse and extend association rules using the non-item dimension to obtain more detailed rules.

With regards to improving the quality of discovered knowledge, several approaches have been mentioned in the introduction. We also note other interesting research that discusses the similarity between patterns to discover really useful patterns [9]

As stated in [2] discovering association rules is an important data mining problem, and there has been considerable research on using association rules in the field of data mining problems. The associations' rules algorithm is used mainly to determine the relationships between items or features that occur synchronously in the database. For instance, if people who buy item X also buy item Y, there is a relationship between item X and item Y, and this information is useful for decision makers. Therefore, the main purpose of implementing the association rules algorithm is to find synchronous relationships by analyzing the random data and to use these relationships as a reference during decision making [2]. One of the most important problems in modern finance is finding efficient ways to summarize and visualize the stock market data to give individuals or institutions useful information about the market behavior for investment decisions.

The enormous amount of valuable data generated by the stock market has attracted researchers to explore this problem domain using different methodologies. [3] Investigated stock market investment issues on Taiwan stock market using a two stage data mining approach. The first stage Apriori algorithm is a methodology of association rules, which is implemented to mine knowledge and illustrate knowledge patterns and rules in order to propose

stock category association and possible stock category investment collections. Then the K-means algorithm is a methodology of cluster analysis implemented to explore the stock cluster in order to mine stock category clusters for investment information. By doing so, they propose several possible Taiwan stock market portfolio alternatives under different circumstances [3].[13] focuses on employing SVM for identifying evolving trends in stock trading but the approach becomes far complex and near unfeasible for real world implementation .

The most notable work to be cited was [14] where authors tried computer vision based but fully dependent on human perception of Stock trading trends using auditory and visual receptors but nevertheless does not address the issue of orchestrating the process. Till date there has been no comprehensive system for scalable analytics of Intraday Stock Market Trading datasets produced out of heavy intraday transactions

IV. PROPOSED SYSTEM

There have been many systems proposed to find correlated patterns in Market basket analysis. Some systems focused mainly on Stock Market Trend prediction using Market basket analysis techniques.

Our problem is to address the problem of analysis of intraday stock market transaction data to obtain correlated patterns of rise and fall in stocks values. Hence our work focuses on

- Studying the nature of intraday stock transaction data from Google and Yahoo finance and
- Propose , develop A Framework for Correlated Pattern Mining in Intraday Stock Trading Transactions
- Considering the big data nature of data we would attempt to extend the system as scalable in-database analytics.

The proposed work being a more modular oriented approach of system design multiphase system architecture is deemed to be used.

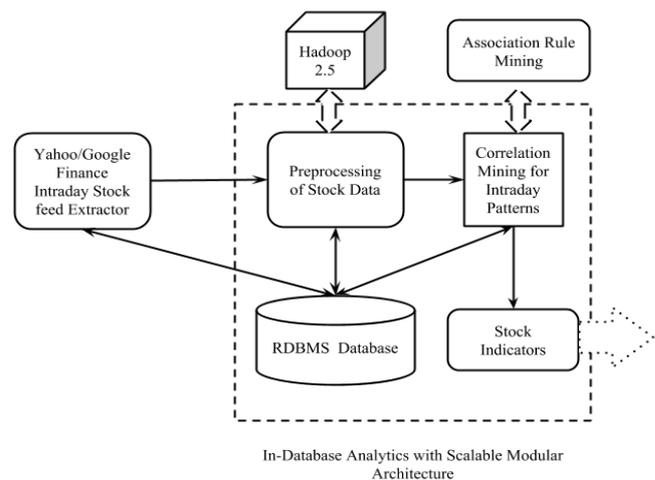
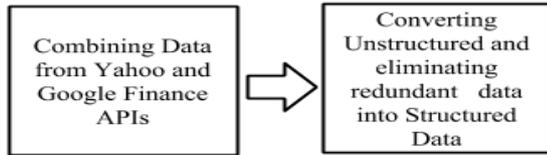


Fig. 1

Initially the datasets of nature of big data have to be obtained to act as foundation data for proof of concept big data analytics of Intraday Stock Market transactions. These datasets will be pulled and scrapped from various web

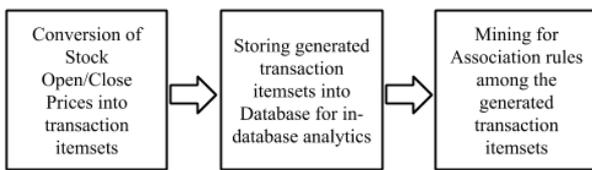
sources, Google and Yahoo Finance being most preferred sources for pulling intraday datasets.



Processes of Preprocessing Module

Fig. 2

The obtained data is processed with help of custom written code scripts. The main objective of processing stage is preparation of datasets for mining associative patterns resembling strong correlations among various stocks with rise and fall of stock values. Since the data being mixture of structured and unstructured a Map Reduce style of processing of datasets could be adopted [11]. The process depicted in Figure 3



Correlation Mining for Intraday trading Patterns

Fig. 3

After pre-processing stage, the processed data will be mapped with proper Relation Database schemas. RDBS would serve the purpose of data store for scalable in-database analytics. Mining of correlated associative patterns of rise and fall of stock in intraday trading is last phase where standard association mining techniques will be employed, without losing the scalability aspect of Modular design. Figure 1 depicts the architectural overview of Modular design of proposed system.

Further detailed tasks of Correlation mining module involves generating, transaction item sets from the processed data. Also the generated transaction item sets have to be further exported to database where proper schemas with optimized indexing strategy ensure quick in-database analytics. The in-database analytics module further mines the transaction item sets for association rules which form the basis of finding correlated trading patterns in intraday trading transactions of stock market. This entire process has been depicted in Figure 2.

The Stock indicator module as shown in Figure 4, on inputs in form of correlated trading patterns as association rules, would be indicating signals as to indicate Buy or Sell signal.

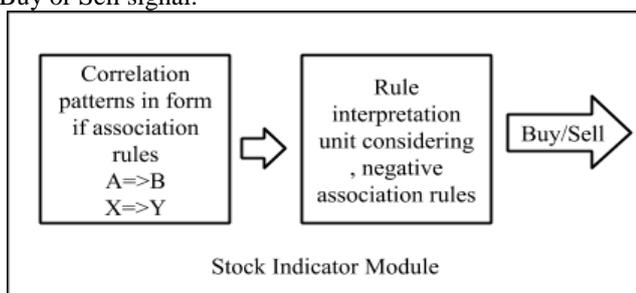


Fig. 4

V. SUMMARY

Modular approach based system design looks feasible and promising in addressing scalability when processing intraday stock market transaction datasets. In-database analytics lets third party vendors use the system without much hassles since most organizations prefer SQL based traditional analytics owing to faster responses, ours one being a SQL based analytics module is a best candidate in such scenarios.

As future extension to this work we propose that this proposed system can be easily integrated in Stock Forecasting Softwares which rely on association mining of inter-transaction association rule mining.

REFERENCES

- [1] Discovery, analysis and presentation of strong rules, Knowledge Discovery in Databases, MIT Press, Cambridge.
- [2] Agrawal R, Imilienski T, Swami A (1993). Mining association rules between sets of items in large databases, In Proceedings of the ACM SIGMOD international conference on management of data.
- [3] Shu-Hsien L, Hsu-hui H, Hui-wen L (2008). Mining stock category association and cluster on Taiwan stock market, Expert Systems with Applications. 500 Index futures prices, J. Bus. Res.
- [4] Mittermayer, M-A., and Gerhard F. Knolmayer. "Newscasts: A news categorization and trading system." Data Mining, 2006. ICDM 06. Sixth International Conference on. IEEE, 2006.
- [5] Hongjun Lu, Ling Feng, and Jiawei Han. 2000. Beyond inter transactional association analysis: mining multi dimensional inter transaction association rules. ACM Trans. Inf. Syst. 18, 4 (October 2000), 423-454. DOI=10.1145/358108.358114 <http://doi.acm.org/10.1145/358108.358114>
- [6] Ke, Yiping, James Cheng, and Wilfred Ng. "Correlated pattern mining in quantitative databases." ACM Transactions on Database Systems (TODS) 33.3 (2008): 14.
- [7] Hellerstein, Joseph M., Christopher Ré, Florian Schoppmann, Daisy Zhe Wang, Eugene Fratkin, Aleksander Gorajek, Kee Siong Ng et al. "The MADlib analytics library: or MAD skills, the SQL." Proceedings of the VLDB Endowment 5, no. 12 (2012): 1700-1711.
- [8] S. Goil and A. Choudhary (1999). A parallel scalable infrastructure for OLAP and data mining. IDEAS, Canada.
- [9] S. Tsumoto and S. Hirano (2003). Visualization of rule's similarity using multidimensional scaling, ICDM, USA.
- [10] Ke, Yiping, James Cheng, and Wilfred Ng. "Correlated pattern mining in quantitative databases." ACM Transactions on Database Systems (TODS) 33.3 (2008): 14.
- [11] Woo, Jongwook, and Yuhang Xu. "Market basket analysis algorithm with Map/Reduce of cloud computing." The 2011 International Conference on

- Parallel and Distributed Processing Techniques and Applications (PDPTA 2011), Las Vegas. 2011.
- [12] Johan Bollen, Huina Mao, "Twitter Mood as a Stock Market Predictor," *Computer*, vol. 44, no. 10, pp. 91-94, Oct. 2011, doi:10.1109/MC.2011.323
- [13] Lean Yu; Huanhuan Chen; Shouyang Wang; Kin Keung Lai, "Evolving Least Squares Support Vector Machines for Stock Market Trend Mining," *Evolutionary Computation, IEEE Transactions on*, vol.13, no.1, pp.87,102, Feb. 2009
- [14] Nesbitt, Keith V., and Stephen Barrass. "Finding trading patterns in stock market data." *Computer Graphics and Applications, IEEE* 24.5 (2004): 45-55.

