Bi-Clustering Approach for Discovering Periodic Pattern from Historical Data

Mr. Manish Sapkal, Prof. V.S.Nandedkar

Department of Computer Engineering, PVPIT, SavitribaiPhule Pune University, Pune, India

ABSTRACT: As financial growth increased, trading system play key role in current running world. Technical analysts always tries to find out price patterns and market trends for making investment in finance markets and these patterns helps to take right decisions related to trading system and making investments. For taking such a decision, someone needs to analyze the price movement and provide trading rules to guide investors, so that they can take correct trading decisions. For analyzing and taking trading decisions, we proposed to develop Biclustering mining to discover effective trading patterns that contain a combination of indicators from historical financial data series. Biclustering is actually a special branch of clustering algorithms because it clusters the data along the row and the column simultaneously in a 2-D data matrix. The mined patterns are considered as trading rules and can be divided in three trading actions (buy, sell, and no-action signals).

KEYWORDS: Data Mining, Bi-clustering Learning; K Nearest Neighborhood Method, Machine Learning.

I. INTRODUCTION

Stock market also called as share market is the aggregation of sellers and buyers of stocks or shares. Stock exchange is place by which people and companies can trade stocks. In current development world, many countries, companies, retailers, manufactures exchanging their goods, stocks or shares. Trade in stock market means transfer for money of a stock or security form seller to buyer. Participants in stock market are either small individual stock investors or large trade investors. Their sell or buy orders may be executed by a stock exchange trades. For making such investment in stock market, Technical analysts try to find out appropriate trading rules in the stock market. They study the historical data, primarily price and volume, to forecast the direction of prices and make trading decisions based on the predictions. But, investors can't find correct results related to trading system, as we all know stock prediction is difficult to predict because its structure is dynamic and highly complicated. As growth changes its price pattern will change. To avoid such a faulty prediction, investors need right trading rules. In this paper, we proposed to develop Bi clustering method. Bi clustering mining is to discover effective technical trading patterns that contain a combination of indicators from historical financial data series. Bi clustering is actually a special branch of clustering algorithms because it clusters the data along the row and the column simultaneously in a 2-D data matrix. The mined patterns are considered as trading rules and can be divided in three trading actions (buy, sell, and no-action signals). These classifications will be useful for making investments and taking right trading decisions in financial markets.

II. BACKGROUND

Historical data ranging from 5-10 years of financial markets forms certain price patterns based on crowd psychology and behavior. These prices patterns factors in all the known factors like socio-economic changes, company's earnings, political factors, natural disaster, internal company matters, company fundamentals etc.
It's practically impossible to study all these parameters about a certain stock or index or any other financial entity in real time scenario. Hence an idea is to first calculate technical indicators and then apply bi-clustering algorithm to identify the patterns for taking decision about buy, sell or no action. As technical chart factor in all the known news, future events, fundamental factors and hence it's not necessary to know all these factors to arrive at trading decision.
Trading rule that is formed based on the outcome of Bi-Clustering algorithm can be used to take decisions and being an automated system, it doesn't take manual efforts and energy to study the stock fundamentals.Stock market (Any other financial entity) is a nonlinear dynamic system that is influenced by a lot of factors such as national policies, the

economic environment, supply-demand relationships, etc. It is not easy for investors to make a correct trading decision at the right time as there is no opportunity to consistently gain profit over the average in the market.

Technical study with various indicatorsand respective patterns is considered as importantsignal for making trading judgments in financial markets. Though, it is veryhard for investors to find useful investing/trading rules built on numerous technical indicators applied on huge historical data and price fluctuations. Hence it is necessary to find a reliable and intelligent method to identify best investment instrument like Equity, commodity, currency etc to beat the inflation and to build wealth with minimal risk.Bi-clustering methods are successfully applied in other domain like medical science (e.g. To identify diabetic samples using historical sample knowledge base). This gives confidence about reliability of this algorithm.

III. LITERATURE SURVEY

In this chapter we review overall literature of system. Reference [1] proposes using 3D subspace clustering systems to mine guidelines that are related to high stock price returns. This method aggregatesfinancial entities that have similar fundamentals (financial ratios) and high price returns across years.

Those classes speak to the choice that would be taken in the framework The outcome will likewise be plotted to an outline for a simpler pattern examination. A few frameworks are constructed in light of budgetary business sector specialized study signs like RSI (Relative Strength Index), Histogram, Simple Moving Average, Money Flow Index, illustrative SAR and so on [2] strategy is orchestrating markers set into choice tree in view of monetary element exchanging/contributing guidelines and order which portray the standards and make purchase, hold, and offer classes which spoke to conclusions in speculation choices. Official choice classes are concentrated on for their plausible addition, arithmetical mean yield, and collective capital list.

Framework [3] portray a shrewd money related substance combining so as to exchange framework bolster vector machine (SVM) calculation and box hypothesis of monetary element. This hypothesis trusts a fruitful stock purchasing/offering for the most part happens when the cost viably breaks out the past wavering box into one all the more new box. In the framework, bolster vector machine calculation is used to make expectations of the new top and base of the swaying box. At that point an exchanging system taking into account the case hypothesis is worked to make exchanging and venture judgments. The different stock development arrangements like bull (Positive), bear (Negative) and turbulent marketplace are utilized to evaluate the attainability of the technique.

The resultant methodology is proposed to be utilized as a judgmental bolster instrument or as a self-sufficient simulated broker if reached out with an interface to the stock trade. Machine learning methodology is likewise exceptionally successful for securities exchange examination. Framework [4] characterize a stock value forecast model will be made utilizing ideas and strategies as a part of specialized examination and machine learning. The subsequent forecast model ought to be locked in as mimicked specialists that can be utilized to choose stocks to exchange on any given stock trade.

In comprehension the troubles confronting speculators amid the venture choice procedure; consider the instance of normal stocks in money related markets that deliver by and large essentially extensive return throughout the years than the sparing record. Be that as it may, a commendable scope of speculators abstain from understanding these substantial returns because of a reality that 'to huntfor high profits traders must acknowledge vast dangers. Model [5] figured as a multicriteria optimizingprototype (gain by the grossyield and decreasing the misfortune that can happen) to be fathomed for the commitment proportion of every investing prediction model segment in the contributing set. Its officially exhibited that the suggested methodology of joining distinctive investingchoiceof methodsresults in obvious surge of increases and additionally critical diminishing in the most extreme misfortune.

Past examination demonstrated encouraging outcomes on the possibility of appropriately forecasting the value course of a stock or market file. Some of that work is said in this study. [6] proposed 5-days-ahead and 10-days-ahead analytical prototypes are readied utilizing the arbitrary timberlands systems. The models are developed on the notable information of the CROBEX and Zagreb Stock Exchange from a few sections. Various specialized signs, which are well known in quantitative examination of money related markets are chosen as model inputs.

A few frameworks characterize bi-grouping developments into two, on the premise of whether the arrangement is characterized on a solitary bunch or numerous bunches. Single bicluster is called as restricted example. Else we call the example a globalized design. [7] address this imperative issue and properlystudy a few bi-grouping calculations as far as the bicluster designs they endeavor to find. We deliberately outline the necessities for natural examples and demonstrate the requirements forced by bi-grouping calculations that characterize their ability to perceive such arrangements. We providepracticaloutcomes from a painstakingly composed proving ground to assess the force of the utilized pursuit procedures.

The framework [8] demonstrate the subtle element clarification of k-means arrangement. [9] subgroups of DNAswhich have similartraits under subdivisions of circumstances, so we say that they express in pair i.e. co express and act autonomously under different subsets of conditions. Finding such coexpressions can be steady to find genomic realities, for example, DNA systems or DNA communications. That is the reason, it is of most extreme essentialness to make a simultaneous gathering of DNAs and circumstances to distinguish gatherings of DNAs that are coexpressed in bunches of circumstances. This kind of collection is called bi-bunching. It likewise centered around bi-bunching of genome expression dataset. The staying of this article is organized as beneath - First, a few definitions identified with bi-grouping of microarray information. We then present in area 3 some assessment capacities and bi-grouping calculations. At long last, we demonstrate to approve biclusters by means of bi-bunching apparatuses on microarrays datasets.

Framework [10] characterize creative philosophy for client division which is the rudimentary sympathy toward an operational CRM (Customer Relationship Management). At first, the chi-square factual study is connected to choose set of qualities and K-implies method is connected to measure the significance of every component. Consequently DBSCAN i.e.(Density based spatial grouping of uses with noise)procedure in view of thickness is introduced to compose the clients into three bunches (First, Second and Third class). At last, bi-bunching taking into account enhanced Apriori calculation is utilized as a part of these 3 gatherings to get more definite data.

IV. PROPOSED SYSTEM

List of technical indicators like RSI, SMA. EMA, ADX, ROC,MACDetc represent the column of table.This data is taken as input to Bi-clustering algorithm. Taking average set of values of output of Algorithm forms a trading rule.Classification of current dataset is done by applying Support Count Method / k-NN neighborhood method and use of trading rule formed in previous step.
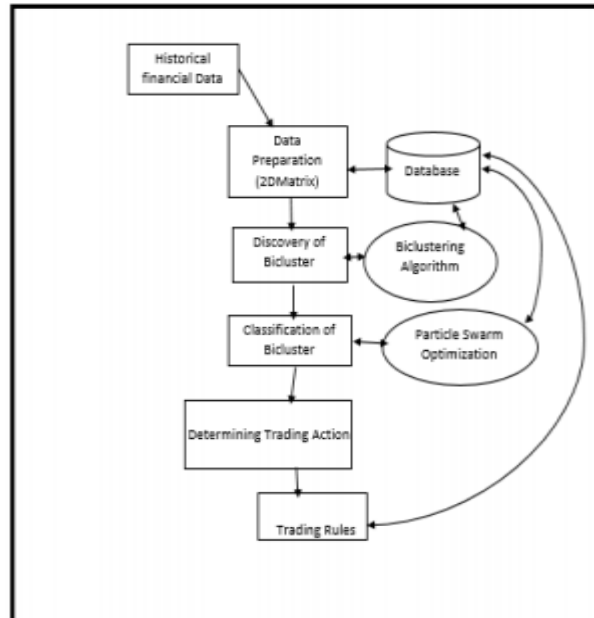
Fig 1: System Architecture of Proposed System

The Proposed work as follows:

Step 1: Input data preparation: For a stock or a financial composite index, a 2 D data matrix is constructed using historical data where the rows consist of trading days and the columns consist of technical indicators with different parameters and future returns.

Step 2: Discovery of Biclusters: After generation of data matrix, we search for bi-clusters which is sub matrix with coherent patterns. For searching bi-cluster we apply Bi-clustering Algorithm. Step 1: To find similar elements in single column for this agglomerative algorithm is applied to each column of data matrix. Step 2: To expand each Bicluster then filter out the duplicate bi-clusters.

Step 3: Classification of the Biclusters: A bicluster is translated into a trading rule by averaging each column. Because each column of a bicluster corresponds to a specific technical indicator. An effective trading rule can help the trader to make a trading decision by predicting a moving direction of the financial price. Three types of trading actions occur in the stock market, i.e., buy, sell, and no action, so all of the trading rules are then classified into three trading sets: 1) buy set 2) sell set and 3) no-action set. Particle swarm is proposed for optimization of best trading actions.

Step 4:Determining of Trading Actions: The trading actions are determined based on the matching of trading rules and trading days. For each trading day, the values of all the technical indicators are calculated. Since a trading rule is a vector of technical indicators with specific values, a trading day is considered to match a trading rule when the values of the corresponding technical indicators of the trading day are very similar to the values of the technical indicators of a trading rule.

Mathematical Model

Identify the input as S = {D, T, C}
Where, D is set of data matrix, T is set of historical technical indicators, C is set current value of technical indicators
Output -

S = {B,Tr }

B = {B | B is set of bi-cluster}
Tr = {Tr | Trading rule }
Identify processes as P,
P= {N(k), T(k), H(L), V(L), C(L),R(L)}
Where,

N(k) = Normalization of input data matrix
T(k)= Calculation of technical indicators
H(L)= Horizontal clustering
V(L)= Vertical clustering
C(L) = Merge the clusters
K(L) = Apply k-nn method to form a trading rule
Mathematical description of Dataset preprocessing and input parameter preparation for Bi-clustering algorithm:
1. Calculate SMA (Simple Moving Average)-10 and SMA-30 using below formula

$$\text{SMA10 or SMA30} = \sum_{k=0}^{n} \frac{(sum = sum + n)}{TotalDays}$$

2. Calculate EMA (Exponential Moving Average)-10 and EMA-30:
EMA = [Todays Close – EMA (priorworking day)] x multiply factor + EMA (priorworkingday).

3. Calculate RSI (Relative Strength Index) -14
RSI14= **[**100– 100/(1 + RS*)];
Where,
RS* = Mean of n days' positivecloses / Mean of n days' negative closes.

4. Calculate ROC (Rate of Change):
ROC = [(close-close n period ago)/(close n period ago * 100)]
Calculate the future return using below formula

$$FRVi = \frac{(CP\_avei - Cpi)}{CPi} * 100$$

Where CP-avei is Average closing price on i'th trading day

5. Normalization of technical indicator

$$Vn(i,j) \frac{(Vo(i,j) - Vmin(j))}{Vmax(j) - Vmin(j)}$$

Where, $Vmax(j)$ is maximum value of j'th indicator, $Vmin(j)$ is minimum value of j'th indicator in datamatrix, i is current row number.

6. Calculate MSRS(Mean Square Residue Score) -To assess the coherence of the elements of a bicluster.

$$H(R,C) = \frac{1}{|R||C|} * \sum_{i \in R j \in C}^{n} (aij - aiC - aRj + aRC)^2$$

$$aiC = \frac{1}{|C|} * \sum_{j \in C}^{n} aij$$

$$aRj = \frac{1}{|R|} * \sum_{i \in R}^{n} aij$$

$$aRC = \frac{1}{|R||C|} * \sum_{i \in R \, j \in C}^{n} aij$$

$\delta$is a homogeneity threshold defining the maximum permissibledivergencein the entities of the bicluster. A submatrix is called a δ bicluster if $H(R, C) \leq \delta$ for some $\delta \geq 0$. The homogeneity threshold is set by users.

Algorithm
Input A Data Matrix X
Output A Set of Bi-Cluster (Bic_Set)
Let HC_Set={ } and Bic_Set = { }
For each column (l) in X

Perform a hierarchical clustering algorithm with a distance threshold T_{hc} on l. We obtain a set of clusters in the single column l, denoted as l_Set.
   Add l_Set into HC_Set.
EndFor

For each cluster(C) in HC_Set
   C' = C;
For each column l
   Add l to C', we obtain C'', Calculate the MSRS for C'', denoted as *lsd.*
Iflsd<Tsd, then C' = C'', **EndIF**
EndFor
   Add C'intoBic_Set.
EndFor
Filter out the duplicate clusters.
Output Bic_Set

V. RESULTS AND DISCUSSION

In the proposed implementation we compare with Kelvin Sim 3D subspace clustering approach [1]. The results get around of satisfactory level. The proposed system accuracy get around 80 to 95%. Here figure 3 shows the comparison our proposed result with existing system [1]. Here X shows the different methods use for creating trading rules and Y shows accuracy level of each method. So, we can conclude even our system reflect estimated results of proposed system is better than existing approaches.
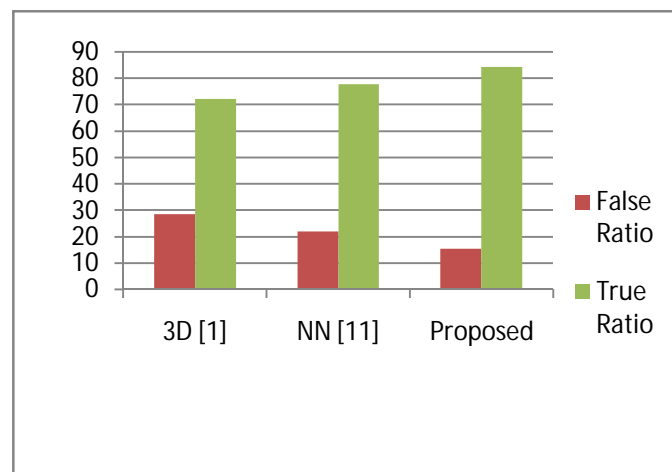

Figure 2: Proposed system analysis with existing approaches.

VI. CONCLUSION

Trading system is having more importance in development of world. Technical analysts need to take right decisions related to trading system for making investments. The Proposed system is developed for analyzing and generating the trading rules. To implement such a system, the proposed Bi-clustering mining is used to generate effective technical Biclusters.The mined patterns are considered as trading rules and can be divided in three trading actions (buy, sell, and

no-action signals).Proposed bi-clustering approach is applied for the optimization of trading actions. By these trading rules, it will be easy for investors to take a right decision related to trading system.

REFERENCES

[1] Kelvin Sim and Vivekanand Gopalkrishnan proposed 3D Subspace Clustering for Value Investing by the IEEE Computer Society in 2014.

[2] F.X. Satriyo D. NugrohoTeguh BharataAdj, SilmiFauziat proposed the DECISION SUPPORT SYSTEM FOR STOCK TRADING USING MULTIPLE INDICATORS DECISION TREE in I st International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE) year 2014.

[3] Qinghua Wen, Zehong Yang, Yixu Song, PeifaJia proposed Intelligent Stock Trading Method established on SVM Algorithm & Oscillation Box Prediction Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009.

[4] Jan Ivar Larsen [4] proposed Predicting Stock Prices Using Technical Analysis and Machine Learning Norwegian University year June 2010.

[5] Tarek M. Gadallah and M. NashatFors proposed New Methodology for Merging Multi Measures Trading Judgment Models, Industrial Engineering and Operations Management Dubai, United Arab Emirates (UAE), March 3 – 5, 2015.

[6] T. Manojlovic and I. Stajduhar proposed Predicting Stock Market Trends Using Random Forests Algorithm - A Example of the Zagreb Stock Exchange MIPRO 2015, 25-29 May 2015, Opatija, Croatia.

[7] DorukBozdag, Ashwin S. Kumar and Umit V. Catalyurek proposed Comparative Analysis of Bi-clustering Algorithms National Cancer Institute Grant R01CA141090; by the U.S. DOE SciDAC Institute Grant DE-FC02-06ER2775; by the USA National Science Foundation in Scholarships 1-CNS- 064396 2-OCI-0904809, 3-OCI-0904802 and CNS-0403342.

[8] KardiTeknomoproposedK-Means Clustering methodology Teknomo, Kardi. K-Means (k-NN) Clustering Lessons."URL-'http:\\people.revoledu.com/kardi/ tutorial/kMean'"

[9] Haifa Ben Saber and MouradElloumi proposed A NEW SURVEY ON BI-CLUSTERING OF MICROARRAY DATA NatarajanMeghanathan et al. (Eds) : NeTCoM, CSIT, GRAPH-HOC, SPTM - 2014

[10] Xiaohui Hu and Haolan Zhang A New Customer Segmentation Framework Based on Bi-clustering Analysis JOURNAL OF SOFTWARE, VOL. 9, NO. 6, JUNE 2014.

[11] FORECASTING OF INDIAN STOCK MARKET INDEX USING ARTIFICIAL NEURAL NETWORK. :Manna Majumder1 , MD Anwar Hussian.