

Bi-Clustering Mining to Discover Effective Technical Trading Patterns

^[1] Mayuri Patil, ^[2] Sarita Patil

^{[1][2]} Department of Computer Engineering, GHRCEM, Wagholi, Pune University, Pune

Abstract – Past database of financial market prices are used to forecast the trend of prospect prices. The developed financial entity technical patterns that contain a permutation of indicators from historical economic data series. Calculation of technical indicators is done based on historical data. The take price forecast system uses a descriptive two-step reasoning approach. This work inventively put forward the use of biclustering mining to determine out patterns are regarded as trading rules and can be categorized as three trading actions (Buy, Sell and No Action) based on support value. K nearest neighborhood (K-NN) procedure (Updated Version) for categorization is applied to trading days in the testing period. Our work offers a practical and efficient algorithm and mathematical system for determining a trading rule which can be used to take informed decision while trading.

Keywords- Data Mining; Bi-clustering Learning; K Nearest Neighborhood Method; Support Count Method; Learning of Trading Rules; Bi-clustering Algorithm; Trend spotting in Financial Markets; Machine Learning.

I. INTRODUCTION

Financial entity fundamental parameters can be calculated using various financial ratios. For example, the return on-equity ratio measures a stock's efficiency in using its assets to generate profit, the debt to equity ratio calculates the portion of the stock's assets which are debts/liabilities. All these fundamental entities are factored in in technical charts and hence study of technical charts is very important to take trading decisions.

The goal of proposed model is to deal with the problem of spotting in trend in financial markets and also represent a novel methodology and mechanism for taking right decisions in trading. In this new approach, proposed work is based on bi-clustering algorithm with support count and K-Nearest Neighborhood Method.

Feature Generation module of our project provides a knowledge-intensive and computationally efficient analysis of past prices which can be analyzed further in a second layer of reasoning.

The field knowledge applied in the module is therefore limited to methods and techniques in technical study. Technical study references include numerous stock analysis techniques, some of which involve complicated and complex price patterns subjective in both discovery and analysis. These methods would be both computationally costly to identify and assess, and have subsequently been disregarded. We thus apply Occam's razor to the choice of procedures in technical analysis, concentrating on the most prevalent indicators that can be

efficiently operationalized and are intuitive in interpretation.

II. PROJECT IDEA

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.

III. MOTIVATION

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 indicator sand respective patterns is considered as important signal for making trading judgments in financial markets. Though, it is very hard 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.

IV. LITERATURE SURVEY

In this chapter we review overall literature of system. Reference [1] propose using 3D subspace clustering systems to mine guidelines that are related to high stock price returns. This method aggregates financial 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 abstains from understanding these substantial returns because of a reality that 'to hunt for high profits traders must acknowledge vast dangers. Model [5] figured as a multi criteria optimizing prototype (gain by the gross yield 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 investing choice of methods results 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 properly study a few bi- calculations as far as the bicluster designs they endeavor to find. We deliberately grouping outline the necessities for natural examples and demonstrate the requirements forced by bi-grouping calculations that characterize their ability to perceive such arrangements. We provide practical outcomes 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 DNAs which have similar traits 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 co expressions 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 co expressed 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 Apriority calculation is utilized as a part of these 3 gatherings to get more definite data.

V. SYSTEM ARCHITECTURE

Proposed Work.

Proposed work consist of series of preprocessing and data clean up calculations and application of Bi-clustering algorithm on cleaned dataset.

As a first step, daily historical data consisting of open, close, high and low price, volume is prepared (This represent rows of a table). Five year's dataset is obtained using NSE's archival facility for financial entities.

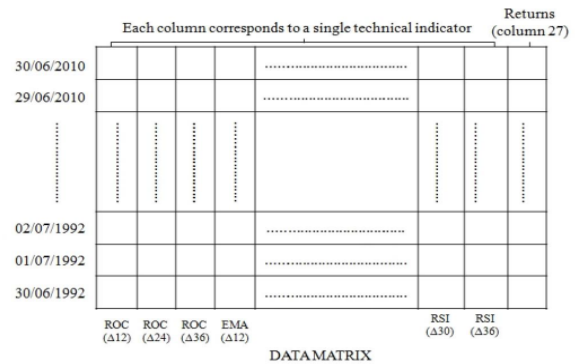


Fig 1: Input Data Matrix

List of technical indicators like RSI, SMA, EMA, ADX, ROC, MACD etc. 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.

Architecture

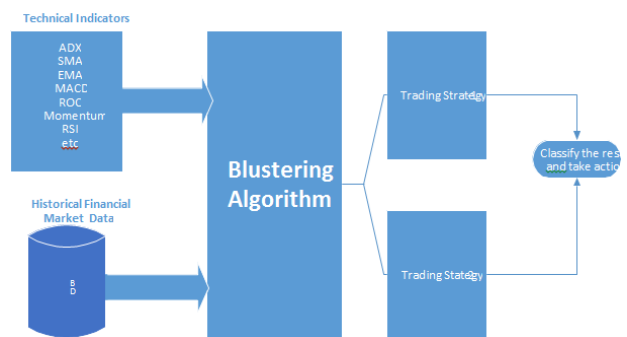


Fig 2: System Architecture of Proposed System

VI. MATHEMATICAL MODELING

Where, D is set of data matrix, T is set of historical technical indicators, C is set current value of technical indicators

Output -

$$S = \{B \text{ Tr}\}$$

$$B = \{B \mid B \text{ is set of bi-cluster}\}$$

$$\text{Tr} = \{\text{Tr} \mid \text{Trading rule}\}$$

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

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

2. Calculate EMA (Exponential Moving Average)-10 and EMA-30:

EMA = [Today's Close - EMA (prior working day)] x multiply factor + EMA (prior working day).

3. Calculate RSI (Relative Strength Index) -14

$$\text{RSI14} = [100 - 100 / (1 + \text{RS}^*)];$$

Where,

RS* = Mean of n days' positive closes / Mean of n days' negative closes.

4. Calculate ROC (Rate of Change):

$$\text{ROC} = [(\text{close} - \text{close } n \text{ period ago}) / (\text{close } n \text{ period ago} * 100)]$$

Calculate the future return using below formula

$$\text{FRVi} = \frac{(\text{CP}_{\text{avei}} - \text{Cpi})}{\text{Cpi}} * 100$$

Where CP_avei is Average closing price on i'th trading day

5. Normalization of technical indicator

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

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.

$$H(R, C) = \frac{1}{|R||C|} * \sum_{i \in R, j \in C} (a_{ij} - a_{iC} - a_{Rj} + a_{RC})^2$$

$$a_{iC} = \frac{1}{|C|} * \sum_{j \in C} a_{ij}$$

$$a_{Rj} = \frac{1}{|R|} * \sum_{i \in R} a_{ij}$$

$$a_{RC} = \frac{1}{|R||C|} * \sum_{i \in R, j \in C} a_{ij}$$

δ is a homogeneity threshold defining the maximum permissible diver genie 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.

VII. 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.

If lsd < Tsd, then $C' = C''$, **EndIf**

EndFor

Add C'intoBic_Set.

EndFor

Filter out the duplicate clusters.

Output Bic_Set

VIII. RESULTS AND DISCUSSION

In proposed system, the experimental results are obtained in the form of precision and comparing predicted and actual values. In this experiment we studied existing systems output over a period of Jan -09 to Mar-09 and Apr 09 to Jun 09. Overall accuracy of existing system is about 70%. However, proposed system has exceeded the expectations and the accuracy is more than 85%. These experimental results are used to demonstrate the usefulness of our suggested methodology by contrasting it with the existing approaches.

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 gets 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.

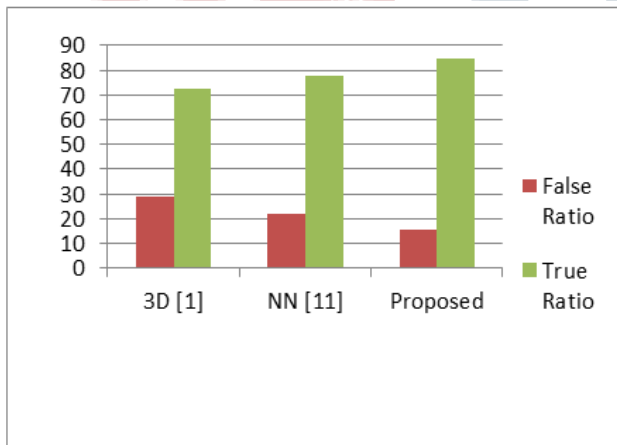


Figure 3: Proposed system analysis with existing approaches.

CONCLUSION

In the proposed system we studied the different bi-clustering approaches as well clustering. We got the idea of proposed approach how to apply such scenarios on

historical data like financial market or any benchmark data. We successfully applied bi-clustering techniques to historical financial data to calculate technical indicators and plot the trend to decide about buy, sell or no action.

All methods exceeded the expectations using real time data and hence a sound performance for trading in financial markets. We expect that the novel idea of converting the bi-clusters into trading rules will lead to a new research topic in finance theory and applications.

REFERENCES

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

[2] F.X. Satriyo D. NugrohoTeguhBharataAdj, 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] DorukBozdog, 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] KardiTeknomoproposed K-Means Clustering methodology Teknomo, Kardi. K-Means (k-NN) Clustering Lessons. "URL-[http:// people. revoledu. com/ kardi / tutorial/ kMean](http://people.revoledu.com/kardi/tutorial/kMean)"

[9] Haifa Ben Saber and MouradElloumi proposed A NEW SURVEY ON BI-CLUSTERING OF MICROARRAY DATA Natarajan Meghanathan 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.

