

Discovering Approach of Classification for Stock Market Prediction using CART with AVL Tree

Garima Saini

Student of Masters in Technology, Smitha Jha

Computer Science, Department of Computer Science

Birla Institute of Technology, Mesra, India Birla Institute of Technology, Mesra, India

Abstract: Data mining techniques have been successfully shown to generate high accuracy of stock price movement. In the field of data mining, the classification techniques are best known for producing accurate, rapid and straight forward results. The goal of classification algorithms is to generate precise and accurate results. In this paper, we have implemented the data mining technique in stock market prediction using CART with AVL tree. We are extracting knowledge from huge raw data which contain several up and down values of stock indexing and it is very difficult, time consuming that exceeds hand capabilities.

Keywords—Stock market, CART, AVL tree

I. INTRODUCTION

A stock market is an open market for the trading of company stock and derivatives at an agreed price; these are securities listed on a stock exchange as well as those only traded privately. The stock market is also called the subordinate market as it involves trading between two investors. Stock market gets investors together to buy and sell shares in companies. Share market arranges prices according to supply and demand.

A stock that is highly in demand will increase in price, whereas as a stock that is being heavily sold will decrease in price. Companies that are permitted to be traded in this market place are called "listed companies" [1].

"The stock market is one of the most significant sources for companies to raise currency." This allows businesses to be openly traded, or raise additional capital for expansion by selling shares of ownership of the company in an open market.

The stock market is often considered the prime indicator of a country's economic strength and development. Rising 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 behaviour of the stock market.

An index is an algebraic composite measure of the movement in the overall market or industry. Basically, indexes allow measuring the performance of a group of companies over a period of time. The movements of the prices in a market or section of a market are captured in price indices called stock market indices.

II. RELATED WORK

Abdullah Al-Luhaib et al in [3] examined to attempt to guess the direction of daily price changes. Back propagation neural network has been applied to predict the direction of price changes for the listed stocks in SSM. In this paper the price change in SSM ranges between -10% and 10%. The target has a representation of three classes 1, -1 and 0 that respectively represent the increase, decrease or insignificant change in the stock prices.

Samarth Agrawal et al in [4] presented an innovative approach for indicating stock market decisions that the investor should take for minimizing the risk involved in making investments. In this paper system uses Adaptive neuro-fuzzy inference System (ANFIS) for taking decisions based on the values of technical indicators. Weighted moving averages, divergence and RSI (Relative Strength Index) technical indicator are used.

Sunisarimchareon et al in [5] presented a classification process of the Stock exchange of Thailand index using adaptive evolution strategies. The prediction process does not require the knowledge of the functional form a priori. In this paper at each recursion step, genetic algorithm is used to evolve the structure of the prediction function, whereas the coefficient is evolved by evolution strategies. In this paper SET index can be reasonably forecasted with only two factors, the Hang Seng index and minimum loan rate. The proposed method in this paper also achieves a lower prediction error when compared with multiple regression method.

III. PROPOSED WORK

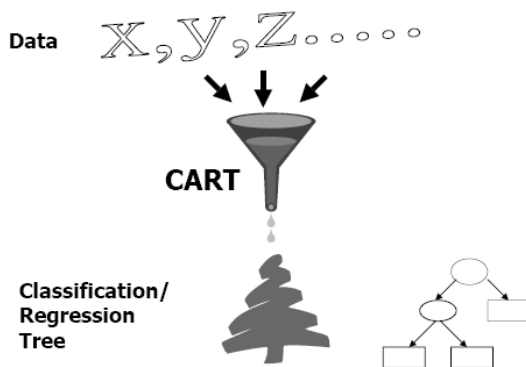
We have implemented CART with AVL tree using C language and Visual Studio for graphics. Machine learning is one of the most advanced concepts in present research scenario. Therefore exploring machine learning along with data mining and its learning algorithms has lots of scope to work. In machine learning and data mining, classification is best for producing correct, quick and straight forward results and hence among several techniques of machine learning, classification has been selected. CART (Classification and Regression Tree) is capable of handling discrete/categorical features and provide quick, true and easy classification results and therefore it is chosen for classification of Indian stock

market data [6]. Firstly we will describe CART algorithm then flow of CART after that insertion of AVL tree algorithm, lastly advantages of CART.

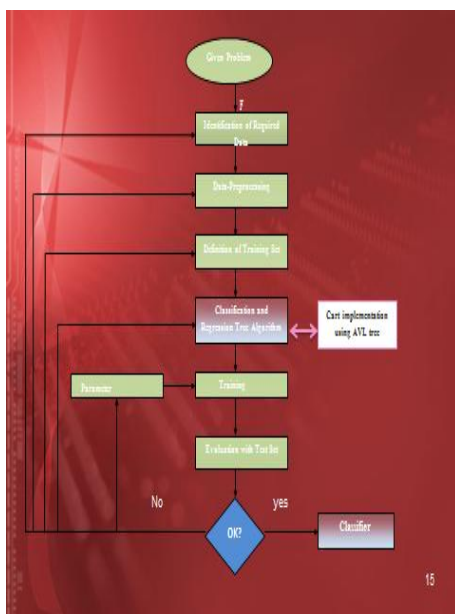
A. CLASSIFICATION AND REGRESSION TREE (CART)

It is developed by Leo Breiman, Jerome Friedman, Richard Olsen and Charles Stone in 1984. Its splitting rules are always couched in the form "An instance goes left if CONDITION, and goes right otherwise, where the CONDITION is expressed as "attribute $X_i \leq C$ " for continuous attributes. In CART algorithm following concept is simply used for making a decision tree [2]. It is a classification method it uses historical data to construct decision trees. Decision tree is then used to classify new data.

In decision tree top most node is called root node and each root node is get subdivided into sub nodes called child node, again each child is treated as parent node as splitting criteria are performed until some stopping criteria is not reached, then terminal node is declared and each nodes are assigned to some classes as in this work terminal node is companies of the stock market dataset [2].



B. FLOW OF CART USING AVL TREE



According to flow chart for CART. First step is problem given then second step is the identification of the required data then next step is data pre-processing. This step is very important for data mining and machine learning algorithms. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training set. Pre-processing is very helpful for knowledge discovery database process. It also improves the performance of classification algorithm. In this dissertation classification and regression tree (CART) algorithm is used for classification. After pre-processing CART algorithm is used. At this step we implement the AVL tree algorithm, we build the AVL tree. This algorithm is based on mainly four steps, first step consists of tree building, during which a tree is built using recursive splitting of nodes and maintain the height of tree. After this step is training. During training parameter tuning is done then evaluation on test data is performed. In final step decision is made if decision is yes then classification is performed by classifier. If no then again it will get switch to identification of required data.

During the CART step, we insert an AVL tree algorithm. At this step we put company name and rate directly then we can obtain tree with maintain height and also obtain line chart, pie chart.

C. INSERTION AVL TREE ALGORITHM

At the fourth step of CART algorithm, we are implemented AVL tree concept.

Insertion AVL Trees

Step 1: Set up the pointers:

1. A - Points to the node (Contains Company Name & Rate) that is out of balance. If more than one node is out of balance then select the one that is furthest from the root. If there is a tie then you missed a previous out-of-balance.

B - Points to the child of A in the direction of the out-of-balance

C - Points to the child of B in the direction of the out-of-balance

F - Points to the parent of A. This is the only pointer of these 4 that is allowed to be NULL.

Step 2: Define the appropriate algorithm:

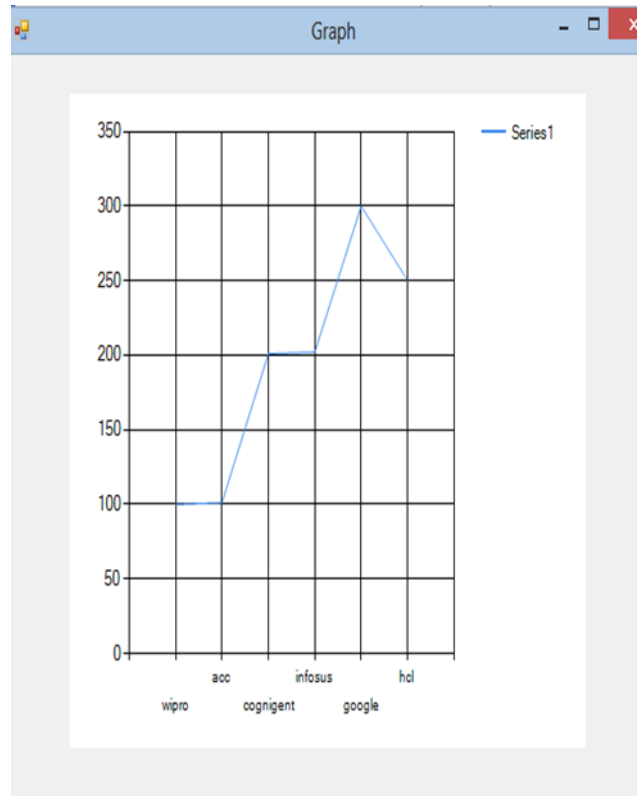
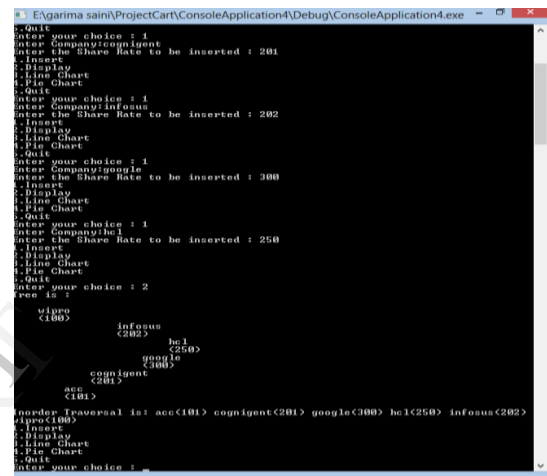
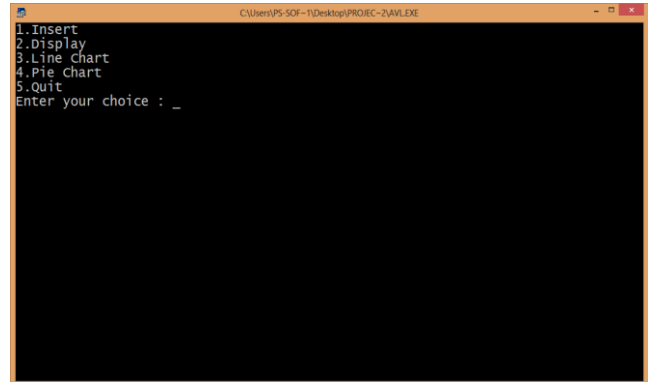
The first letter of the algorithm represents the direction from A to B (either **Right** or **Left**). The second letter represents the direction from B to C (either **Right** or **Left**).

Step 3: Follow the algorithm

III EXPERIMENT RESULT

<p>Right-Left B.Left=C.Right A.Right=C.Left C.Right=B C.Left=A If F=NULL C is the new Root of tree Else If F.Right=A F.Right=C Else F.Left=c End if End if</p>	<p>Left-Right A.Left=C.Right B.Right=C.Left C.Left=B C.Right=A If F=NULL C is the new Root of tree Else If F.Right=A F.Right=C Else F.Left=C End if End if</p>
--	--

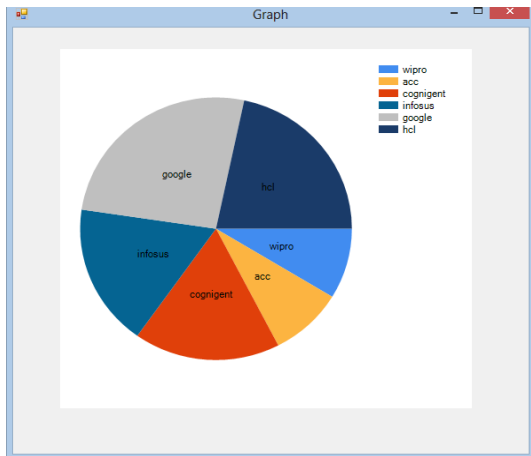
<p>Right-Right A.Right=B.Left B.Left=A If F= NULL B is the new Root of tree Else if F.Right=A F.Right=B Else F.Left=B End if End if</p>	<p>Left-Left A.Left = B.Right B.Right=A If F= NULL B is the new Root of tree Else if F.Right=A F.Right=B Else F.Left=B End if End if</p>
---	--



Using this AVL tree algorithm we developed a code in C and Visual studio for line chart and pie chart. So the market analyst and the financial analyst can easily understand the status of stock market. The generated code will give precise and accurate results

D.ADVANTAGE OF CART WITH AVL TREE

- In CART, AVL tree was far quicker.
- In CART ,AVL has Average behavior but better than random binary search tree
- Insertion times in the AVL tree were dramatically quicker than in the binary search tree.
- In CART, AVL tree allowed for more than 30000 items to be inserted, whereas the generic binary searchtree would result in a stack overflow.
- It results are invariant to monotone transformations of its independent variables..
- It can handle datasets with difficult structures.



IV.CONCLUSION

The previous researches have been done only on foreign stock market data therefore in this work; CART is used with Indian stock market data for classification purpose because results obtained during classification are easy to interpret due to its tree structure. Experimental results have shown that CART is better implementation using AVL tree algorithm. It is accurate, fast and precise.

REFERENCES

1. D. VenugopalSetty, T.M.Rangaswamy, K.N.Subramanya "A Review on Data Mining Applications to the Performance of Stock Marketing", International Journal of Computer Applications, Volume 1 – No. 3, pp-24-34, 2010.
2. Chee Jen Chang, "Partitioning groups using classification and regression tree in biomedical research", 2002.
3. Abdullah Al-LuhaibKhaled Al-GhoneimYousefAl-Ohali, "Dynamic targets for stock market prediction", International Conference on Signal Processing and Communications ICSPC, 2007.
4. Samarth Agrawal, Manoj Jindal and G. N. Pillai, "Momentum Analysis based stock market prediction using adaptive neuro-fuzzy inference system (ANFIS)", Proceeding of the international multiconferences of engineering and computer scientist, IMECS, 2010.
5. SunisaRimcharoen, DarichaSutivong and PrabhasChongstitvatana, " Prediction of the stock exchange of Thailand using adaptive evolution strategies", Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05),IEEE, 2005.
6. Thair Nu Phyu, "Survey of classification technique in data mining", International multiconference of engineering and computer scientists, 2009.