International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-1, January 2016
ISSN: 2395-3470
www.ijseas.com

IMPLEMENTING CLASSIFICATION FOR INDIAN STOCK MARKET USING CART ALGORITHM WITH B+ TREE

Kalpna Singh[1], U.Datta[1] ,K.K.Joshi[1]
[1]Deptt. of Computer Science & Engg., MPCT Gwalior. India

Abstract:

So Many Researchers always try to work on Data Mining to find out something new to enhance the efficiency of searching & decision making analysis. Indian Stock market is the good area where we do some research and Indian stock market data has always been a certain appeal for researchers. Classification and regression tree (CART) , Quadratic discriminate analysis (QDA) and linear discriminate analysis (LDA) are introduce for classification of Indian stock market data. Some researchers implemented algorithm using Decision tree & AVL Tree. AVL tree Work Better than Decision Tree. I am implementing the CART algorithm using B+Tree to enhance the performance because smaller misclassification reveals that CART algorithm using B+Tree performs better classification for Indian stock market data as compared to LDA and QDA algorithms.

Keyword:LDA,QDA,CART,AVL, B+TREE.

1.Introduction

CLASSIFICATION AND REGRESSION TREE (CART*)*

CART technique was developed by Leo Breiman, Jerome Friedman, Richard Olsen and Charles Stone in 1984. CART is one of the best methods of building decision trees in the machine learning community CART create a binary decision tree by splitting the values/rate at each node, according to a function of single attribute. CART uses the gini index to find out the best split. CART follows the following principle of constructing the decision tree. Binary tree that all elements in the left sub-tree of a node n are less than the contents of n, and all elements in the right sub-tree of n are greater than or equal to the contents of n. Binary tree is useful data structure when two-way decisions must be made at each point in a process. For example search the all duplicates values in a list of numbers. One way doing this is each number compare with given node and precede it. However, it involves a large number of comparisons. The number of comparisons can be reduced by using a binary tree. in this work terminal node is companies of the stock market dataset [2*].The major disadvantage of a binary search tree is that the height of tree must be maximum N-1. This means that the time required to perform deletion as well as insertion and for the many other operations should be O(N) in the worst case . if one can want a tree with a minimum height .Thus, our goal is to keep the height of a binary search tree O(log N) .This can be accomplished* when we implement CART Using AVL tree algorithm.

Why CART using B+ Tree is useful? Using AVL Tree we found that its useful in memory based search but in case of disk based search B+ tree allow storage of large amounts of data on disk, indexed by key, with records accessible in O(*log n*) time where *n* is the number of nodes in the tree.

Here I am implementing stock market analysis for intraday trade and find out the least price ,max price and average price movement according to time, this helps client to find out the nature of particular share in market.

1. Literature Review

[5] Implementation of CART Using AVL TREES

- The major drawback of a binary search tree is that its height can be as large as N-1
- This means that the time required to implement insertion of node and deletion of node and many other operations of node can be O(N) in the worst case
- If we want a tree with a minimum height
- If we want a binary tree with N nodes having height of least O(log N)
- Thus, if our main aim is to maintain the height of a binary search tree O(log N)
- These trees are called balanced binary search trees. Examples are Splay trees, AVL trees, and B+ trees

In binary search tree, optimization of searching is totally dependent on the order of inserted element . If binary search tree is complete balanced tree then this optimization can be achieved . We know that complete balanced tree is an ideal situation but we can be near to this optimization of search time for insertion and deletion the Russian mathematics G.M Adel's son . Vel'skii and E.M Landies came with new technique for efficient for balancing binary search tree called AVL tree on their names . This tree has technique for efficient search and insertion, so AVL tree behaves near to complete binary search tree.

So now our purpose is to maintain the binary search tree as complete binary search tree. In binary search tree it is not possible to maintain the same height of left and right subtree of any nodes as in complete binary tree . But with AVL tree we can get the binary search tree height of left and right subtree will be with maximum difference 1 which is close to complete binary search tree.So we can define AVL tree as –
An AVL tree is a binary search tree where height of left and right subtree of any node will be with maximum difference 1.
Each node of AVL tree has a balance factor . Balance factor of a node is defined as the difference between the height of left subtree and right subtree of a node.
Balance factor= height of left subtree – height of right subtree

A node is called right heavy or right high if height of its right subtree is one more than height of its left subtree .A node is called left heavy or left high if height of its left subtree is one more than height of its right subtree . A node is called balanced if the heights of right and left subtree are equal .

The balance factor will be 1 for left high,-1 for right high and 0 for balanced node .So in an AVL tree each node can have can have only three values of balance factor which are -1, 0, 1 or we can say the absolute value of balance factor should be less than or equal to 1.

1. Left to left rotation—insertion in left subtree of left child of pivot node .
2. Right to right rotation—insertion in right subtree of right child pivot node.

3. Right to left rotation—insertion in left subtee of of right child of pivot node.
4. Left to right rotation—insertion in right subtree of left child of pivot node.

Code for constructing a node for B+ tree

```
struct node{
 char name[30];
  long rate;
  int n; /* n < M No. of keys in node will always less than order of B tree */
  int keys[M-1];
 struct node *p[M];   /* (n+1 pointers will be in use) */
}*root=NULL;
```

The balance factor of a node is $h_L - h_R$ where h_L is the height of its left subtree and h_R is the height of its right subtree. As soon as a node's balance factor becomes +2 or –2, it is rebalanced and balance factor becomes 0.

Disadvantage of AVL

The disadvantages is the complex rotations used by the insertion and removal algorithms needed to maintain the tree's balance. Avl Tree is good for searching node when data is stored in Memory. But in case of data searching from disk this technique is quite slow.

3. Related Work

[1]Sumit Garg et al In his research he describe that single algorithm may not help to find the suitable data in data mining. His paper focuses on comparative analysis of various data mining techniques and algorithms.

[2] Shashikumar G. Totad , Geeta R. B. et al In this research paper they discuss about the issues that is being carried out on parallel and distributed data mining. They find out some core data mining algorithms such as decision trees, discovery of frequent patterns, clustering, etc., to find quick result for parallel processing in data mining.

[3]Mohd Mahmood Ali1, Lakshmi Rajamani et al In this paper they proposed a data classification method using AVL trees enhances the quality, accuracy and stability of data mining problems. But this research is useful only in memory based

Searching. But when used this technique in disk based serach the this technique slow down the process.

[4] Sneha Soni Samrat et al. In this paper, they made combination of three supervised machine learning algorithms, classification and regression tree (CART) , linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are proposed for classification of Indian stock market data, which gives simple interpretation of stock market data in the form of binary tree, linear surface and quadratic surface respectively. But as we discuss earlier binary tree & AVL tree is best when we uses this techniques in memory based problems.

[5] Garima Saini et al In her paper she found that the classification techniques are best known for producing accurate, rapid and straight forward results. She implemented data mining technique using CART with AVL tree to find out the stock market prediction.

4. Proposed Work

We'll implement CART with b+tree using C Language and Visual Studio for graphics.

Machine learning is one of the most advanced concepts in present research scenario. Therefore there are so many techniques for machine learning process along with data mining and its learning algorithms and there are 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 choose a CART (Classification and Regression Tree) which 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 .Firstly we will describe CART algorithm then
insertion of AVL tree algorithm, lastly insertion and advantages of B+ tree.

Research design problem and Formulation

In CART, AVL tree was far quicker then decision tree and has Average behavior but better than random binary search tree but the Insertion times in the AVL tree were dramatically quicker than in the binary search tree.
AVL tree allowed for more than 30000 items to be inserted, whereas the generic binary search tree would result in a stack over flow. Insertion and deletion in AVL tree was very complex and may some
time gives very obscure data instead the real values. Though AVL tree retrieve data from the memory only. And it search the data in its leaf and root node which increases search space and time.

RESEARCH FORMULATION
We will be implementing B+ Tree on Indian stock market so that the data can be retrieve from the disk as well as from memory. As B+Tree searches the value only on the root node which reduces the searching time and space.

5. Conclusion

In this research we will use Data Mining to find out something new to enhance the efficiency of searching & decision making analysis. Indian Stock market is the good area where we will do some research and Indian stock market data has always been a certain appeal for researchers. Classification and regression tree (CART) , quadratic discriminate analysis (QDA) and Linear discriminate analysis (LDA)are proposed for summarization of Indian stock market data. Some researchers implemented algorithm using Decision tree & AVL Tree. AVL tree Work Better than Decision Tree. We will be implementing the CART algorithm using B+Tree to enhance the performance because smaller misclassification reveals that CART algorithm using B+Tree performs better classification for Indian stock market data as compared LDA & QDA algorithms.

References

1].Sumit Garg
Comparative Analysis of Data Mining Techniques on Educational Dataset.
M.Tech Scholar Dept. of Computer Science Shekhawati Engineering College Dundlod, Rajasthan, India
2].Shashikumar G. Totad , Geeta R. B. , Chennupati R Prasanna , N Krishna Santhosh , PVGD Prasad Reddy
Scaling Data Mining Algorithms to Large and Distributed Datasets
3].Mohd Mahmood Ali, Lakshmi Rajamani
Decision Tree Induction: Data Classification using Height-Balanced Tree
Assistant Professor, Dept of CSE, Muffakham Jah College of Engineering & Technology, Hyderabad, India, Professor, Depart-

ment of CSE, University College of engineering, Osmania University, India
4]. Sneha Soni
Classification of Indian Stock Market Data Using Machine Learning Algorithms
Samrat Ashok Technological Institute VIDISHA, M.P., India
5]. Garima Saini
Discovering Approach of Classification for stock market prediction using Cart with AVL
Birla Institute of Technology Mesra India.