

PROGRAMMING FAULT PREDICTION MISTREATMENTQUAD TREE-BASED K-MEANS AGGLOMERATION ALGORITHM

*K. Yuvaraj*¹

ABSTRACT

In this venture a Quad Tree desire expansion calculation and K-Means calculation has been connected for foreseeing issues in the product modules. The point of this venture is twofold. To begin with, K-Means estimation is associated for watching the fundamental gathering centers to be commitment to the Quad Tree. An input threshold parameter delta governs the quantity of initial cluster centers and by variable delta the user will generate desired initial cluster centers. The possibility of agglomeration pick up has been wont to check the standard of groups for examination of the Quad Tree-based organization algorithmic program when contrasted with elective arrangement methods. The clusters obtained by K-Means algorithm were found to have most gain. Second, the Quad Tree-based calculation is connected for foreseeing shortcomings at interims the information. The general blunder rates of this forecast approach unit contrasted with different existing calculations and unit observed to be higher.

Keywords : Clusters, K-means clustering algorithm, EM algorithm, Jvm, java, quad tree, qdk

I. INTRODUCTION

This research paper focuses on clustering and falls under the domain of explorative data mining. It is a

Assistant Professor, Department of CS, CA & IT Karpagam Academy of Higher Education

typical system for measurable information examination. Its use is seen in multi different fields, including machine learning, design acknowledgment, picture examination, data recovery, and bioinformatics. Unsupervised strategies like bunching may be used for fault prediction. This research aims to predict faults in the classification of datasets(ref 1). In this venture a Quad Tree based EM calculation has been connected for foreseeing shortcomings in the arrangement of datasets. The general blunder rates of this expectation approach are contrasted with other existing calculations, for example, K-Means and are found to be better in most of the cases. This research focuses on clustering by partition based method namely K-Means algorithm and model based method namely, EM algorithm.

A definitive objective is to enhance the precision of blame expectation by utilizing EM grouping calculation. It is an expansion of the K-Means grouping calculation. Notwithstanding bunching in light of limiting intra group separate, likelihood is figured for every blend of information point and group. Clustering is done based on a weighted relationship thus derived. A significant decrease in error rates is observed through the proposed system implementation.

2. Problem Definition

This research aims to predict faults in the classification of datasets. The current framework utilizes K-implies

bunching calculation where the exactness of the blame is not precisely anticipated(ref 3). So the proposed EM algorithm is efficient to predict the accuracy of the fault.

3. PROJECT DESCRIPTION

Module Description

1. K-Means Algorithm
2. Quad Tree Based Expectation Maximization Algorithm

K-Means Algorithm

K-Means is an unsupervised clustering method where observations are iteratively relocated among a set of clusters until the convergence criterion is met. This popular algorithm follows a partition dividing. It by then uses iterative development system that undertakings to improve isolating by moving articles beginning with one assembling then onto the following. K-Means gathering is direct, brisk and comprehensively used approach to manage describes or cluster data(ref2).

Mechanism Done In Simple K-Means Clustering

- " Cluster numbers to be choose as k.
- o Randomly produce k groups and decide the bunch focuses, or specifically create k arbitrary focuses as group focuses.
- o Assign each point to the closest group focus, where "closest" is characterized regarding one of the separation measures examined previously.
- o Re-register the new group focuses.
- o Repeat the two past strides until the point when

some union measure is met (as a rule that the task hasn't changed).

Quad Tree Based Expectation Maximization Algorithm

The basic bundle centers are found using a quad tree based figuring. A quad tree is a tree information structure in which each interior hub has precisely four youngsters. Quad trees are regularly used to segment a two dimensional space by subdividing it into four quadrants or locales. The locales might be square or rectangular. Desire Maximization is a kind of model based grouping technique. It endeavors to upgrade the fit between the given picture and some numerical model. Such techniques are regularly in viewed with the suspicion that the information are produced by a blend of hidden likelihood circulations. The EM calculation is an expansion of the K-Means algorithm. The proposed EM calculation is known to be a suitable streamlining for discovering smaller bunches. EM ensures an exquisite merging(ref 4) EM estimation consigns an inquiry a gathering as showed by a weight addressing the probability of enlistment. EM then iteratively resource the things and updates the evaluations. The mistake rate for K-Means calculation and EM calculation are processed, indicating the quantity of effectively and mistakenly arranged specimens by every calculation. Result includes charts showing up on a relative start the amplexness of EM computation with quad tree for fault gauge over the present Quad Tree based K-Means (QDK) show.

Mechanism Done In Quad Tree Based Expectation Maximization Algorithm

First, instate the parameters to some irregular esteems.

Compute the best an incentive for given parameter esteems.

Then, utilize the just-processed esteems to register a superior gauge for the parameters.

Iterate steps 2 and 3 until joining.

4. SYSTEM DESIGN

In K-Means, select the text file that contain pixel values of an image and obtain initial value, centroids and distance value from midpoint to other values (as shown in fig.1.)

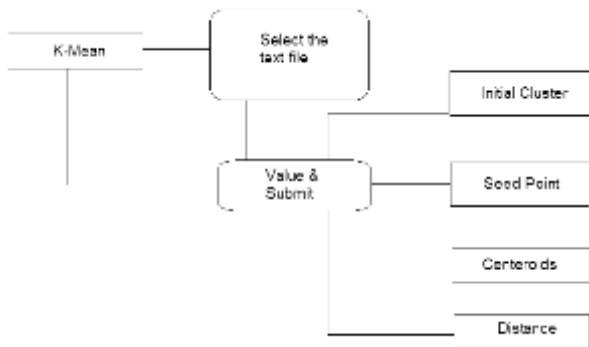


Fig 1: Block Diagram For K-Means

In Quad Tree a particular area is selected in an image, starting and end point is selected by the user on the selected area. Quad tree is used to find the minimum distance and predict the fault in the image (as shown in fig.2)

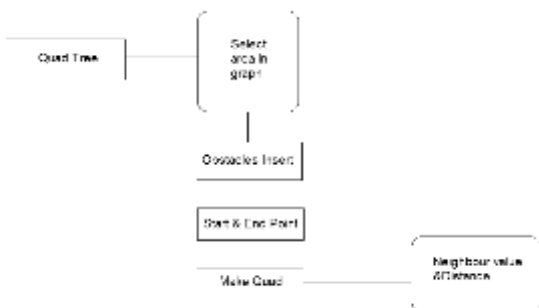


Fig 2: Block Diagram for Quad Tree

Input Design

Input design involves the selection of the best strategy for getting data into the system at the right time and as accurately as possible. The use of well-defined documents can encourage users to record data accurately without omission(ref 5).

The goal of designing input data is to make data entry an easy operation. An efficient input designing will avoid the frequent occurrence of errors. The user interacts with the computer system using the input design.

Objectives of Input Design

Wrong information is the most widely recognized reason for mistakes in information handling, blunders found at the information section can be controlled by the best possible information plan. The information configuration was composed with the accompanying destinations. To create a savvy strategy. To create a savvy strategy and get most abnormal amount of exactness there are various input screens: Selection Panel

Selection Panel includes K-Means, Quad tree EM and chart. User can select any one of the choice. (Refer Fig 9.1).

K-Means Process

K-Means Process has the text file to be chosen and the number of clusters to be made.

Sample Input to be given in text file

(2, 10) (2, 5) (8, 4) (5, 8) (7, 5) (6, 4) (1, 2) (4, 9), (4, 1), (5, 1)

Number of clusters to be made can be given by the user dynamically.

Quad tree EM

Selects the particular area in the image and give starting to end point. Construct the Quad tree to find the minimum distance from midpoint to each pixel value.

Output Design

These rules are followed for both paper and screen yields. Yield configuration is regularly talked about before different parts of plan on the grounds that, from the customer's perspective, the yield is the framework.

K-Means Process

K-Means process forms displays number clusters, minimum value, maximum value and centroids value.

Quad tree EM

Quad tree EM displays the minimum value, neighbor pixel value and distance of each pixel from center point.

Chart

Showcases a diagram demonstrating the examination of blunder rate between k-implies calculation and Quad tree EM calculation(ref6).

Test Plan

A test arranges is a record itemizing a deliberate way to deal with testing a framework, for example, a machine or programming. The course of action conventionally contains a point by point perception of what the inescapable work process.

Table1: Test Plan

TEST CASE ID	TEST OBJECTIVE	STEPS TO BE GIVEN	EXPECTED OUTPUT	TEST RESULT
TP01	To check it is K-Means or Expectation Maximization	Select any one option	It navigates to the selected Algorithm page	Pass
TP02	To check that user have selected K-Means then File Chooser file is Opened	Mandatory select the Text File	Accepts the given text file	Pass
TP03	To check K-Means process	Give the no of partition required	Accepts the input	Pass
TP04	To check K-Means process is executed		Shortest distance value is acquired	Pass
TP05	To check that user have selected EM-quad tree	Select the area	Accepts the input	Pass
TP06	To process the Expectation Maximization Algorithm to find shortest distance		Shortest distance value is acquired	Pass

Test Case Id	Test Objective	Steps To Be Given	Expected Output	Test Result
KMP01	To check that user have selected K-Means then File Chooser file is Opened	Mandatory select the Text File	Accepts the given text file	Pass
KMP02	To check that user have selected K-Means then File Chooser file is Opened	Without selecting the Text File	It does not work	Pass
KMP03	To check number of lines and word	-	Displays the word count and line count from the given text file	Pass
KMP04	To check K-Means process	Give the no of partition required	Accepts the input	Pass
KMP05	To check K-Means process is executed	-	Shortest distance value is acquired	Pass

Table 2: Test Case for K-Means Process

Test Case Id	Test Objective	Steps To Be Given	Expected Output	Test Result
EM01	To check that user have selected EM-quad tree	Select the area	Accepts the input	Pass
EM02	To give start and end points in a selected graph area	Enter the values for start and end point	Accepts the input	
EM03	To process the Expectation Maximization Algorithm to find shortest distance		Shortest distance value is acquired	Pass

Table 3: Test Case for Quad Tree EM-Means Process

5. Results and Discussions :

Comparing both the tables results, it shows that the error finding technique quad tree based EM is better than K-means. The program is tested individually at the time of development using the data and has verified that this program linked together properly(ref9).

The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly.

6. Conclusion

This research reviews the issues with exploitation easy K-Means within the classification of datasets. The viability of Quad Tree essentially based EM agglomeration algorithmic run in foreseeing deficiencies though grouping a dataset, when contrasted with various existing calculations similar to, K-Means has been assessed. (ref 7) The Quad Tree approach assigns acceptable initial cluster centers and eliminates the outliers. K-Means is taken as one of the best strategies to cluster information. However, the planned EM algorithmic rule is employed to cluster information effectively. Consolidating the Quad Tree approach conjointly the EM algorithmic control offers an agglomeration system that not exclusively fits the data higher inside the groups anyway additionally attempts to make them minimal and a considerable measure of significant abuse EM adjacent to Quad Tree makes the order technique faster. With K-means, convergence isn't secured; however, EM guarantees elegant convergence(ref 8). The planned system obtains the suitable initial cluster centers through Quad Tree. These centroids work contribution to the EM

algorithmic control, so expanding the potential outcomes of finding the most straightforward groups. the general error rates of the planned system are found similar to different existing approaches.

7. Future Work

An extension of this project would be to use a headquarters Tree primarily based EM cluster model. The home office tree is utilized as a substitution to the typical Quad Tree approach in this way on get even extra exact bunch focuses/centroids. A home office tree might be a D-dimensional simple of a quad tree. every hub of a home office tree is identified with a bouncing hyper box and each non leaf hub has second adolescents. so base camp Trees region unit anticipated that would yield higher bunch focuses when contrasted with the Quad Tree approach.

Book References

1. Meenakshi PC, Meenu S, Mithra M, Leela Rani P, "Fault Prediction using Quad Tree and Expectation Maximization Algorithm", International Journal of Applied Information Systems (IJAIS) Foundation of Computer Science FCS, Vol 2, No.4, Pages 36-40 May 2012.
2. P.S.Bishnu and V. Bhattacharjee, "Software Fault prediction using Quad tree based K-Means method", IEEE transactions on Knowledge and Data Engineering, Vol-2. No.99, May 2011.
3. Swati M.Varade and Prof.M.D.Ingle, "Overview of Software Fault Prediction using Clustering Approaches and Tree Data Structure" The International Journal of Engineering And Science (IJES), Vol-1, Pages 239-242, Year-2012.

4. McKenna, Y. Raja, and S. Gong, "Tracking color objects using adaptive mixture models", *Image and Vision Computing*, vol. 17, Pages 225-231, 1999.
5. J. Han and M. Kamber, *Data Mining Concepts and Techniques*, second Ed, Morgan Kaufmann Publishers, 2007.
6. E.W. Forgy "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". *Biometrics*, 1965.
7. Y.Ohta, T.Kanade and T.Sakai, Color information for region segmentation *Computer Graphics Image Process.*, 1980.
8. Raphael Finkel and J.L. Bentley. "Quad Trees: A Data Structure for Retrieval on Composite Keys". *Acta Informatics*, 1974.
9. Shneier, M. Calculations of geometric properties using quad trees. *Computer Graphics & Image Processing, Algorithms for computing geometric properties of binary images represented as quad trees*, 1981.