

Fault Prediction Using K-means And Apriori Algorithm

K. Yuvaraj¹ T. Genish²

ABSTRACT

In this venture Quad Tree desire expansion calculation and K-Means calculation have been connected for foreseeing issues in the product modules. The point of this venture is twofold. To begin with, K-Means count is associated for watching the fundamental gathering centers to be commitment to the Quad Tree. An info limit parameter delta administers the amount of introductory bunch habitats and by factor delta the client will create the required starting group communities. The possibility of agglomeration gain has been used to confirm the standard of bunches for the examination of Quad Tree-based configuration algorithmic program when contrasted with elective arrangement methods. The groups got by K-Means algorithmic program were found to possess most addition. Apriori count the general error paces of this gauge approach unit stood out from various existing computations and unit saw as higher.

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

1. INTRODUCTION

This project is entitled as FAULT PREDICTION USING K-MEANS AND APRIORI ALGORITHM Clustering falls under the domain of explorative data mining. It is a typical system for the examination of measurable information. It's use is seen in different fields, including AI, structure affirmation, picture assessment, information recuperation, and bioinformatics. Unsupervised strategies like bunching may be used for fault prediction. This project aims at predicting faults in the classification of datasets. The general error rates of this expectation approach are contrasted with

other existing calculations, as, K-Means and are found to be better in most of the cases. This project centers around bunching by segment-based technique in particular K-Means calculation.

A definite 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 implementation of proposed system.

In This Paper

Quad trees are connected for discovering introductory bunch habitats for K-Means calculation. Client can produce fancied number of bunch focuses that can be utilized as contribution to the straightforward K-Means calculation.

The centroids acquired by the quad tree information structure are contribution to the EM algorithm to yield better results.

Results are shown via charts indicating better throughput using the proposed system

2. SYSTEM ANALYSIS

To predict faults in the classification of datasets, many clustering methods exist to partition a dataset. The existing system use K-implies bunching calculation where the error rate of the faults not precisely anticipated. So the proposed algorithm is efficient to predict the fault.

Existing System

K-Means is a straightforward and prominent approach that is generally used to bunch/order information. In any case, K-Means does not generally ensure best bunching because of

¹Research Scholar, Dept. of CS, CA & IT
Karpagam Academy of Higher Education, Coimbatore

²Research Guide, Dept. of CS, CA & IT
Karpagam Academy of Higher Education, Coimbatore

shifted reasons.

Disadvantages of K-Means

There is difficulty in looking at the nature of the bunches delivered. K-Means regularly requires that the bunches be of circular shape. It doesn't wipe out outliers. It does not guarantee the best portrayal of information in the bunches.

The Proposed System

The proposed calculation is known to be a suitable improvement for discovering smaller groups. The error rate for K-Means calculation is registered, indicating the quantity of effectively and erroneously arranged specimens by every calculation.

Advantages of the Proposed System

Will work regardless of restricted memory (RAM) better throughput with cut down screw up paces of request Reduction in time and computational intricacy.

3. PROJECT DESCRIPTION

Module Description

1. K-Means Algorithm
2. Apriori

K-Means Algorithm

CK-Means is an unsupervised clustering method where observations are iteratively relocated among in set of clusters until the intermingling model is met. This famous calculation follows a segment separating. It by then uses iterative improvement framework that endeavors to update isolating by moving articles starting with one gathering then onto the accompanying. K-Means gathering is direct, fast and extensively-used approach to manage describe or bundle data.

Mechanism Done in Simple K-Means Clustering

- Cluster numbers to choose, k.
- Randomly produce k groups and decides the bunch

focus, or specifically creates k arbitrary focus as group focus.

- Assigns each point to the closest group focus, where "closest" is characterized regarding one of the separation measures examined previously.
- Re-registers the new group focus.
- Repeats the two past strides until the point when some union measure is met (as a rule that task hasn't changed).

QuadTree Based Apriori Algorithm

The Apriori calculation learns affiliation administrations, and is applied to a database containing countless exchanges. Affiliation rule learning is an information digging method for learning connections and relations among factors in a database. By applying the Apriori calculation, we can gain proficiency with the basic food item that are bought together under a affiliation rules. The error rate for K-Means calculation and Apriori calculation are processed, indicating the quantity of effectively and mistakenly arranged specimens by every calculation. Result contains diagrams showing up on a relative reason the sufficiency of Apriori figuring with quad tree for accuse estimate over the present Quad Tree based K-Means (QDK) show.

Mechanism Done In Quad Tree Based Expectation Maximization Algorithm

1. The first is the size of your thing set. Would you like to see designs for a 2-itemset, 3-itemset, and so on.?
2. The second is your help or the quantity of exchanges containing the thing set isolated by the absolute number of exchanges. A thing set that meets the help is known as a successive thing set.
3. The third is your conviction or the unexpected probability of something given you have certain various things in your thing set.

4. SYSTEM DESIGN

Block Diagram for K-Means

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.

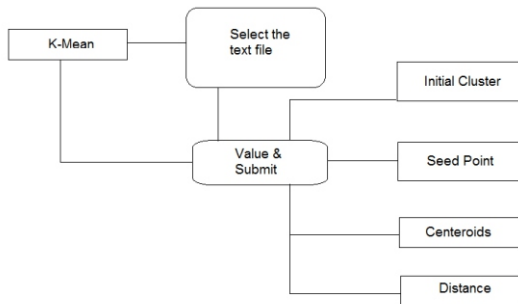


Figure 1:Block Diagram For K-Means

Block Diagram for Quad Tree

In Quad Tree a particular area is selected in an image, start and end points are selected by the user on the selected area. Quad tree is used to find the minimum distance and predict the fault in the image.

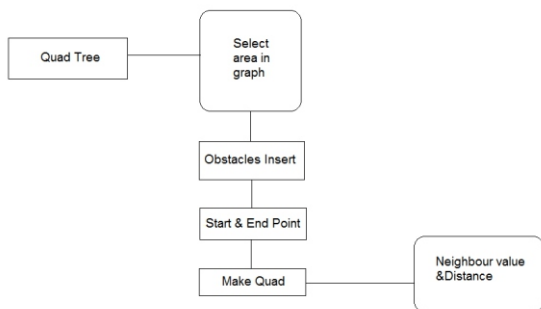


Fig.2 Block Diagram for Quad Tree

Input Design

Information configuration includes determination of the best procedure for getting information into the framework at the perfect time and as precisely as could be expected under the circumstances. The utilization of much characterized archives can urge clients to record information precisely without oversight.

The objective of structuring input information is to make information passage a simple activity. A productive info structuring will maintain a strategic distance from the regular occurrence of mistakes. The client communicates with the PC framework utilizing the information structure.

Objectives of Input Design

Wrong information is the most widely recognized reason for

mistakes in information handling, and errors found at the information section can be controlled by the best possible information plan. The information configuration is composed with the accompanying destinations. To create a savvy strategy. get most abnormal amount of exactness and various input screens are:

Selection Panel

Selection Panel includes K-Means, Quad tree EM and chart. Users can select any one of the choices.

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 Apriori

Selects the particular area in the image and gives start to end points: construct the Quad tree to find the minimum distance from midpoint to each pixel value.

Output Design

These rules are connected with 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 form displays number clusters, minimum value, maximum value and centroids value.

Quad tree Apriori

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

Chart

Showcases a diagram demonstrating the examination of error rate between k-implies calculation and Quad tree EM calculation.

Test Plan

A test orchestrate is a record organizing a conscious method to manage testing a system, for instance, a machine or programming. The arrangement ordinarily.

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

Table Test plan

contains a point by point comprehension of the inevitable work process.

Test Cases

An experiment in programming is a rendezvous of conditions or factors beneath which associate instrument can decide if associate application or programming framework is functioning effectively. The component for choosing if an item program or system has passed or failed. partner explore is generally a singular walk, or sporadically a gathering of adventures, to check the right direct/helpfulness, portions of partner application partner test is a meeting of check inputs, execution conditions, and expected results made for a specific objective, as an example to follow a specific program way or ensure consistence with a particular need. The motivation behind associate experiment is to acknowledge and convey conditions that may be in check. Experiment info is created with the specific goal of deciding if the framework can prepare it effectively. Experiments are

necessary to substantiate effective and satisfactory execution of the item wants.

Test Case Id	Test Objective	Steps To Be Given	Expected Output	Test Result
SP01	To check it is K-Means or Expectation Maximization or graph	Select any one option	It navigates to the selected Algorithm page	Pass

Table Test Case for Selection Panel

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 Test Case for Quad tree APriori

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 Test Case for K-Means Process

5. SYSTEM IMPLEMENTATION

A pivotal face inside the framework life cycle inside the triumphant execution of the new framework style. an essential component of a system assessment work is to shape sure that the new style is actualized to set up standards. Use used here is to prescribes that the method for changing a crisp out of the crate new or a refreshed structure style into accomplice degree movement one change is one part of execution.

Use is that the stage inside the endeavor wherever the speculative style is transformed into a working system and is giving assurance on the new structure for the usages that it'll work with profitability and feasibly. It includes cautious planning, examination of the present framework and its constraints on implementation, style of ways to attain the transmutation, associate degree analysis, of amendment over ways.

- Testing the developed software with sample data.
- Debugging of any errors if identified.
- Creating the files of the system with actual data.

- Making necessary changes to the system to find out errors.

- Training of our personnel.

Apart from planning major task of preparing the implementation are education and training of users. The more complex system being implemented, the more involved will be the system analysis and the design effort required just for implementation. On implementation coordinating committee based on policies of individual organization has been appointed.

Implementation

The program is tested individually at the time of development using the data and has verified that this program linked together properly.

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.

The execution procedure starts with setting up the arrangement for the usage for the framework. As indicated by this arrangement, the exercises are to be done, conversation made with respect to the hardware and assets and the extra gear as to be obtained to actualize the new framework.

The usage is the last and significant stage. The most basic stage in accomplishing fruitful new framework and in giving the client certainty that the new framework will work and be viable. The framework can be executed simply after intensive testing is done and on the off chance that it found to working as per the particular.

This technique likewise offers the best security since the old framework can dominate if the mistakes are found or failure to deal with specific sort of exchange while utilizing new framework.

Steps to be implemented

1. Install the software Net Beans 6.0
2. Install Java development toolkit 6 in the libraries of the system.
3. Include Jfreechart jar file in the bin directory of the java, it is used to display the chart.

6. CONCLUSION

This paper reviews the issues with exploitation easy K-Means within the classification of datasets. The effectiveness of Quad Tree primarily based APriori agglomeration algorithmic rule in predicting faults whereas classifying a dataset, as compared to different existing algorithms comparable to, K-Means has been evaluated. The Quad Tree approach does out adequate starting group habitats and disposes of the anomalies. K-Means is considered to be one of the best procedures to bunch data. The arranged Apriori algorithmic standard is utilized to bunch data viably. Solidifying the Quad Tree approach conjointly the Apriori algorithmic rule offers an agglomeration technique that not solely fits the information higher inside the gatherings in any case moreover endeavors to make them decreased and a lot of significant . misuse Apriori near Quad Tree makes the portrayal strategy speedier. With K-infers, mix isn't ensured about at any rate APriori guarantees rich mixing The organized system gets the proper early on pack networks through Quad Tree. These centroids work commitment to the Apriori algorithmic standard, so extending the potential results of finding the most clear gatherings. the general misstep paces of the masterminded system are found like particular existing philosophies.

Future Work

A development of this endeavor is use a headquarters Tree basically based EM bundle model. The home office tree is used as a substitution to the normal Quad Tree approach right now get even extra precise gathering networks/centroids. A headquarters tree may be a D-dimensional basic of a quad tree. each center point of a focal station tree is related to a

bouncing hyper box and each non leaf center has second young people. so headquarters Trees an area unit expected to yield higher gathering territories when diverged from the Quad Tree approach.

7. 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 (IJ AIS) 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.