A Novel Density based K-means Clustering for Test Case Prioritization in Regression Testing Result-II

[1]Dr. Amit Verma, [2]Rohit Bajaj, [3]Ishadeep Kaur Luthra
[1,2,3]Dept. of CSE, Chandigarh Engineering College, Landran (Mohali), India

Abstract
In this paper, we work on improving the test case prioritization on the basis of clustering approach. A novel density based k-means clustering approach is used to make clusters of different test cases on the basis of statement coverage. Then, prim's algorithm is used to find out the minimum path between different test cases according to their coverage information. Test cases are select from every cluster; which have maximum coverage information. According to Prim's algorithm, we will find the tree of test cases; this technique reduces the test cases numbers. Only those test cases are selected which have maximum coverage information. It will reduce the effort, cost and time also.

Keywords
Test Case Prioritization, Density based Kmeans, Regression Testing

I. Introduction

Software development has come into an era of 40ies. Since then testing has become a major part of the software development. In this competitive world providing the best software with latest technology has become the need of software engineers. To achieve this goal, engineers try to develop new software which is more users friendly. Before implementing this software it goes through number of tests. Sometimes the software does not give the desired result; in that case some changes are required into the developing code. Then the repeated test performed on the updated code is the Regression Testing. The various techniques used in Regression Testing are:

- Retest All
- Regression Test Section
- Prioritization of Test Cases

The first case is very time consuming process and cost inefficient as the tester has to perform the test cases repeatedly. In case of selecting the test section, developers has to be carefull enough that it pick up only relevant test cases and should not waste time on irrelevant test cases. For example: The test cases in which changes have been made and should not pick the test cases which are already being tested. To solve the problem of test section, Prioritization of test cases need to be done. To prioritize the test cases many algorithms have been produced till date such as Greedy view, Particle Swarm Optimization, Spanning Tree algorithm and many more.

Clustering algorithm has been produced to prioritize the test cases, but we have used the concept of novel density based k-means clustering to make clusters of different test cases on the basis of statement coverage. Then, prim's algorithm is used to find out the minimum path between different test cases according to their coverage information. Test cases are select from every cluster; which have maximum coverage information. According to Prim's algorithm, we will find the tree of test cases; this technique reduces the test cases numbers. Only those test cases are selected which have maximum coverage information. It will reduce the effort, cost and time also.

II. Problem Formulation

A The problem of test case prioritization has gained significant attention over the last few years as software testing forms a major section of the whole software development process. The cost of software development is directly dependent on the testing effort. This thesis aims to reduce this cost by prioritizing test cases and running the tests for the selective test cases as per the available time and manpower. There are a number of test cases available which can consume a lot of time and effort. A selective number of test cases needs to be selected which would be otherwise used for the same purpose. The priority of the test cases needs to be decided on the basis of several parameters. The parameter for the test case prioritization needs to be chosen and a model needs to be developed which would set priority among the test cases. First of all a data set needs to be generated which would be utilized for our proposed algorithm testing. Then the dataset needs to be preprocessed for outlier removal and redundancy removal. Then a technique for clustering of the test cases needs to be developed which would be utilized for the above mentioned problem.

III. Methodologies

Steps of Density Based K- Means Algorithm:

Let X = {x1, x2, x3... xn} be the set of data points in Dataset D, Euclidean "ε" (eps).
K is the number of clusters to be found, minPts is a minimum number of neighbors required in ε neighborhood to form a cluster and N is a set of points in ε neighborhood.

Step 1: Start with arbitrary starting point (not visited).
Step 2: Find all the neighbor points of starting points using Euclidean distance "eps".
Step 3: If number of neighbors 'N' is greater than or equal to minPts, then starting point and its neighbors are added to cluster. Starting point is marked as visited; otherwise this point is marked as noise.
Step 4: A new unvisited point is recall and processed for further make a part of cluster or noise.
Step 5: Repeat Step 2, until all points are marked as visited.
Step 6: We have 'm' cluster for each detected clusters; then find cluster centers, 'Cm' by taking the mean find the total number of points in each clusters.
Step 7: If 'm' cluster is greater than 'K' clusters, then join two or more cluster based on density and no. of points to find the new cluster center. Repeat Step 7, until achieving K clusters with 'Ck' centers.
Step 8: Otherwise if 'm' is greater than or equal to the number of clusters initially found by density based clustering algorithm 'l'; select a cluster based on density and number of points split it using k-means clustering algorithm. Repeat Step 8, until achieving k clusters with 'Ck' centers.
Step 9: Apply iteration of k-mean clustering with k and new 'Ck' centers as the initial parameters and label all the clusters with k labels.
Step 10: End

Flow chart of Density Based K-means Clustering Algorithm:

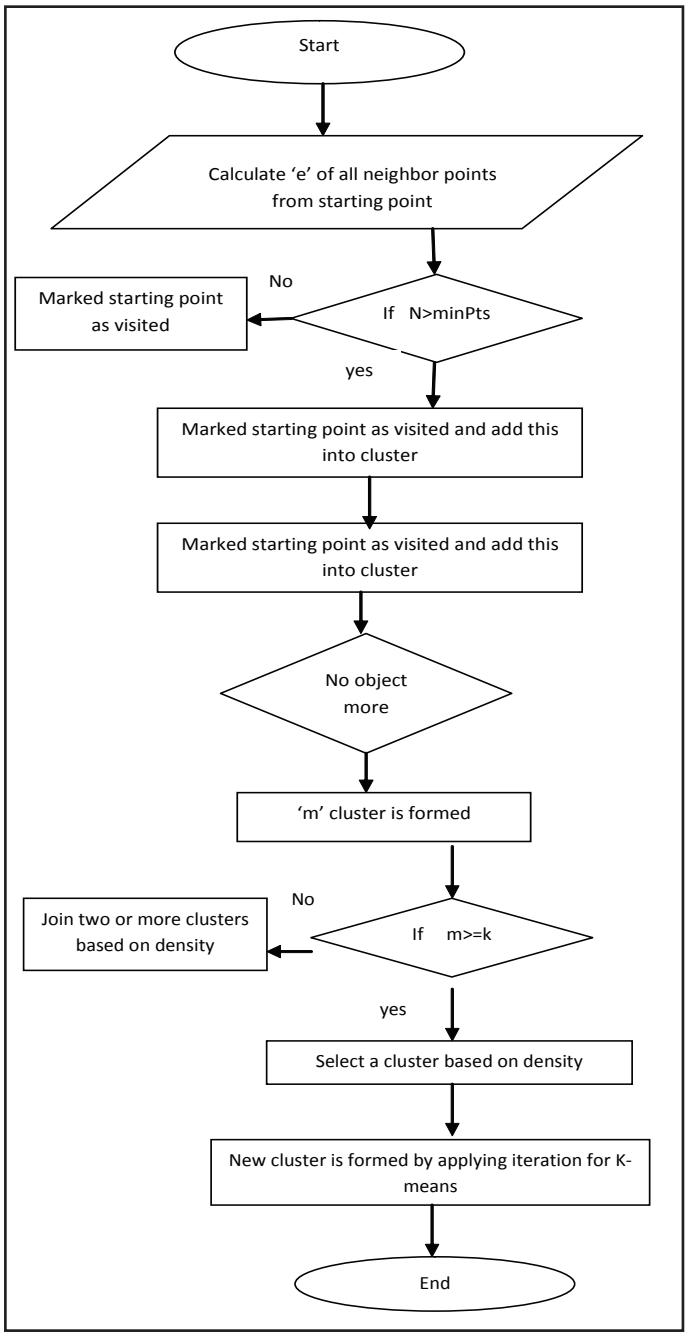

Fig. 1: Flow chart of Density Based K-means Clustering Algorithm

IV. Results

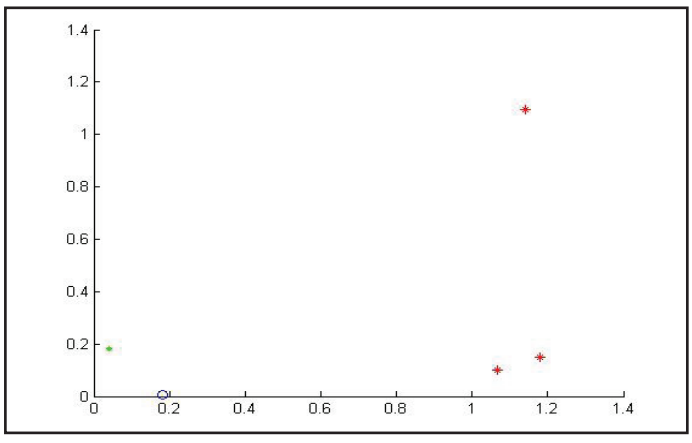

Fig. 2: Clusters Obtained from First Iteration

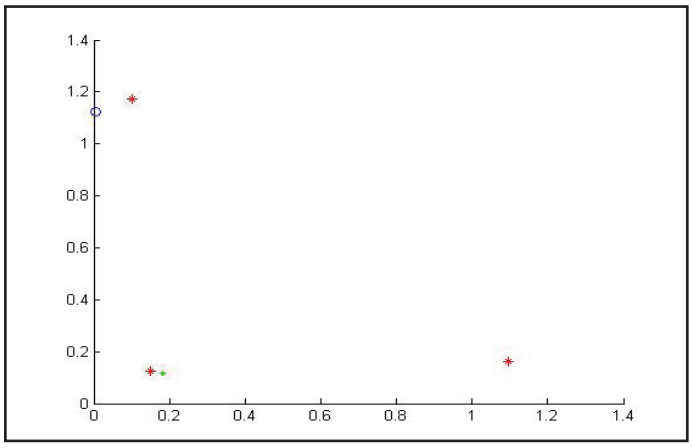

Fig. 3: Clusters Obtained from Second Iteration

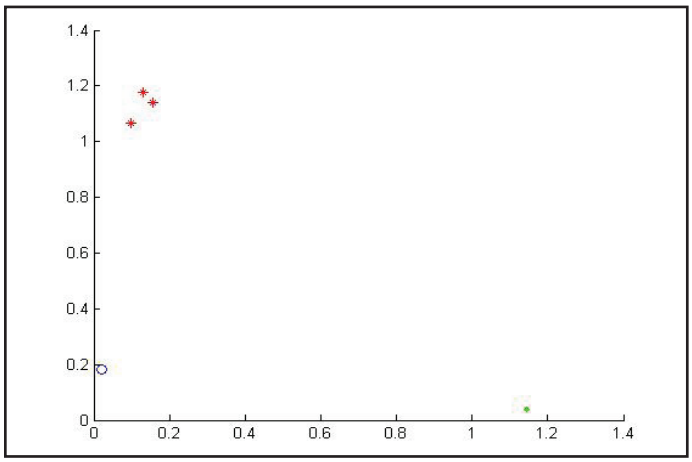

Fig. 4: Clusters Obtained after Number of Iterations

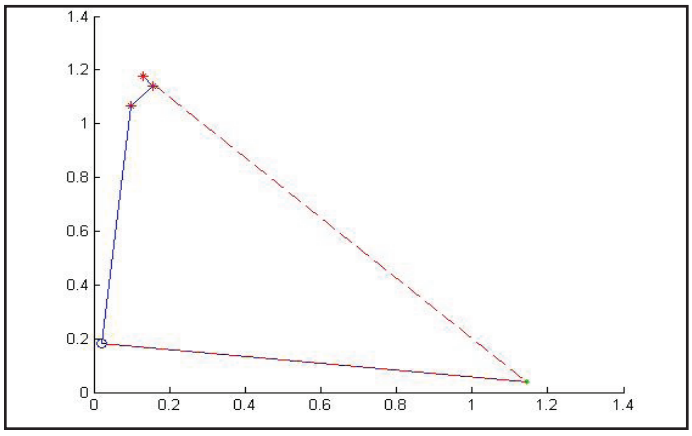

Fig. 5: Representing the Total and Sub-List of Test Cases

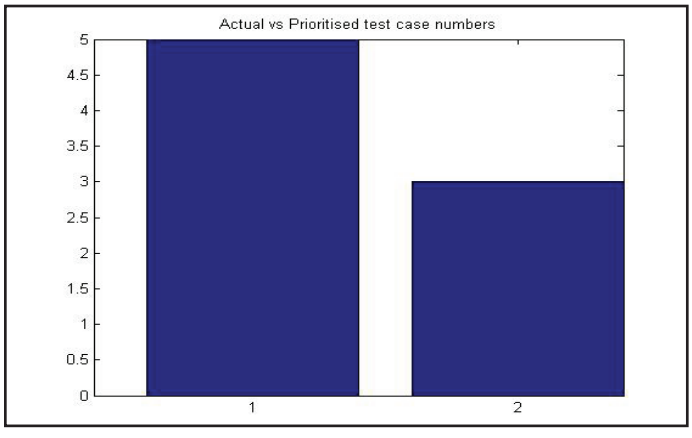

Fig. 6: Actual Vs Prioritized Test Case Numbers

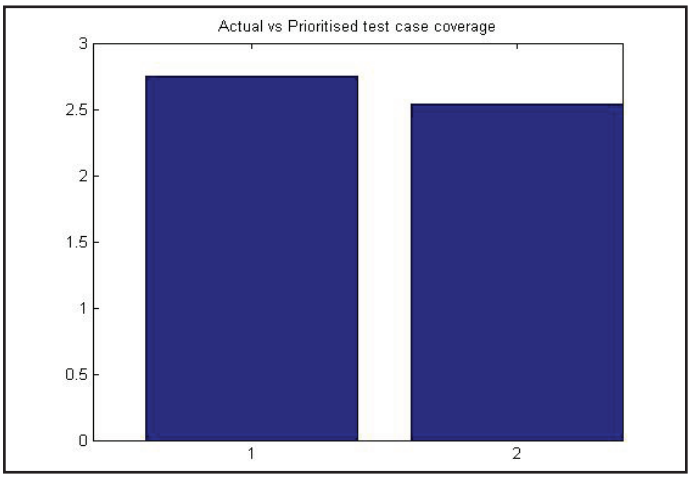

Fig. 7: Actual Vs Prioritized Test Case Coverage

VI. Conclusion and Future Scope

A novel Density based K-means clustering has been applied for software testing. Test Case prioritization has been done using a hybrid algorithm of minimum spanning tree and density based k-means clustering. The results have been found to be quite better than their traditional counterparts. In future other algorithms can be implemented for the same and compared for performance.

References

[1] G. Antoniol, M. D. Penta, M. Harman,"Search-based techniques applied to optimization of project planning for a massive maintenance project", In Proc. of the 21st ICSM, Washington, DC, USA, 2005, pp. 240-249.

[2] Y. Shi, R.C. Eberhart,"A modified particle swarm optimizer," Proceedings of the IEEE World Congress on Computational Intelligence, pp. 69-73, 1 998

[3] Manika Tyagi, Sonika Malhotra,"Test Case Prioritization using Multi Objective Particle Swarm Optimizer", 2014 IEEE

[4] Xiang Chen, Zhaofei Tan, Jian Xia, Pengfei He, "Optimizing Test Case Execution Schedule using Classifiers", Journal of Software, Vol. 9, No. 10, October 2014.

[5] Ryan Carlson, Hyunsook Do, Anne Denton,"A Clustering Approach to Improving Test Case Prioritization: An Industrial Case Study", 2011 27th IEEE International Conference on Software Maintenance (ICSM) 47.

[6] R.Krishnamoorthi, S.A.Sahaaya Arul Mary,"Regression Test Suite Prioritization using Genetic Algorithms", Department of Computer Science and Engineering, Bharathidasan Institute of Technology Anna University, Trichy-24, India

[7] M. R. Garey, D. S. Johnson,"Computers and Intractability: A Guide to the Theory of NP-Completeness", W. H. Freeman & Co., New York, NY, USA, 1979.

[8] G. Rothermel, R. J. Untch, C. Chu,"Prioritizing test cases for regression testing", IEEE Transaction on Software. Eng., Vol. 27, Issue 10, 2001, pp. 929-948.

[9] P. Chu, J. Beasley,"A genetic algorithm for the multidimensional knapsack problem. Journal of Heuristics", Vol. 4, Issue 1, 1998 , pp. 63-86.

[10] R. P. Pargas, M. J. Harrold, R. R. Peck,"Test-data generation using genetic algorithms", Software Testing, Verification and Reliability, Vol. 9, Issue 4, 1999, pp. 183-282.

[11] D. Fatiregun, M. Harman, R. M. Hierons,"Evolving transformation sequences using genetic algorithms", In Proc. of 4th SCAM, 2004 , pp. 66-75.

[12] G. Antoniol, M. D. Penta, M. Harman,"Search-based techniques applied to optimization of project planning for a massive maintenance project", In Proc. of the 21st ICSM, Washington, DC, USA, 2005, pp. 240-249.

[13] Thillaikarasi Muthusamy, Dr. Seetharaman.K, "Effectiveness of Test Case Prioritization Techniques based on Regression Testing", International Journal of Software Engineering & Applications (IJSEA), Vol. 5, No. 6, November 2014.

[15] K. Mumtaz, Dr. K. Duraiswamy,"A Novel Density based improved k-means Clustering Algorithm – Dbkmeans", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 02, pp. 213-218, 2010.

[16] Sebastian Elbaum et al,"Prioritizing Test Cases for Regression Testing", Technical Report # TR-UNL-CSE-2000-0005, Also presented in International Symposium of Software Testing and Analysis, 102-112, August 2000.

[18] Bharti Suri, Shweta Singhal,"Analyzing Test Case Selection & Prioritization using ACO", ACM SIGSOFT Software Engineering Notes, Vol. 36, No. 6, November 2011.