An Efficient Approach for Test Suite Reduction using Density based Clustering Technique

Rashi Chauhan
Dept. of Computer Science & Engg
M.Tech. Student of Amity University
Haryana, India

Pooja Batra
Dept. of Computer Science & Engg
Lecturer, Amity University
Haryana, India

Sarika Chaudhary
Dept. of Computer Science & Engg
Lecturer, Amity University
Haryana, India

ABSTRACT
Density based clustering technique groups similar data objects based on density. In this paper, a methodology has been proposed based on density based clustering technique in order to remove redundant test cases so that time wasted in testing unnecessary test cases can be reduced. Experiments show that how our technique can significantly reduced test suite.

Keywords
Test case, Test suite, Clustering, Selenium, Weka , Redundant

1. INTRODUCTION
Software testing[2] is the process of executing a program in order to find faults that helps the software developers to improve quality of software product and reduce the cost produced by these faults. [2]Software testing means generation of test cases to test correct functionality of software. [2]The output of test case may be either pass or fail. Test suite reduction means eliminating unnecessary test cases which does not contribute to fault detection ability. Software testing is time consuming as it consumes much time in testing redundant test cases. So, our approach deals with elimination of these redundant test cases to save time spent in testing. For this, Density based clustering technique has been used as it is more efficient technique because it overcomes certain limitations of other clustering techniques.

This paper is organized as follows. In section 2, a lot of work done related to area of test suite reduction by others has been discussed. In section 3, a discussion is made on DBSCAN clustering algorithm and tools used for testing and mining test cases. In section 4, research methodology is presented. In section 5, experimental results are shown. In section 6, performance has been evaluated. In section 7, concluding remarks and future work to be done in this area has been discussed. In section 8, acknowledgements have been given. In section 9, references are listed.

2. RELATED WORK
Lilly Ramesh and G.V. Uma [1] discussed a new approach for automatic generation of test cases from software requirements specification. [2]Classification rules are generated first with the help of Weka classifier that classifies SRS into functional and non-functional requirements from which state diagram is derived which is then transformed into test cases upon which clustering techniques are applied to mine test cases.
Kartheek Muthyala et.al. [3] proposed a new technique based on clustering by which test suite can be significantly reduced. By implementing given data set in Weka using K-means clustering and then eliminating redundant test cases by applying pickupCluster algorithm, they provide reduced test suite which again tested for coverage and yielded good results.

Lilly Ramesh and G.V. Uma [2] proposed an algorithm that groups the test cases based on similarity of their execution profiles and samples some representatives to form the reduced test suite. [2]Firstly, various metrics are calculated from test suite and based on results, appropriate metric is selected to cluster test cases and from each cluster sample test cases are selected to reduce test suit size.

Ying Huang and Lu Lu [4] proposed matrix based user session data reduction technique to reduce test suite. Firstly, user session data is gathered and then, service profile is constructed with the help of reverse engineering tools. After this, user session matrix is constructed using these two fields from which reduced data set is selected using concept analysis.

3. BACKGROUND
3.1 DBSCAN clustering
DBSCAN is designed to find non-spherical shaped clusters.[3] It finds the objects that have dense neighbourhoods and connect them with their neighbourhoods to form clusters[1].

Procedure[1]:

- Initially all the objects are marked unvisited.
- Then, randomly select an unvisited object m and mark it as visited .
- Check if its neighbourhood has no minpoints objects,then mark m as noise point.
- Else
- Form a new cluster C and add m to C.
- Add remaining objects of neighbourhood of m to set N.
- Repeat same procedure for each unvisited objects of N until all objects are visited.

3.2 Selenium
Selenium is freeware automation tool for testing web based applications. It performs functional regression testing. Selenium IDE helps in recording test cases. It uses Javascript enabled web browser.

3.3 Weka
Weka is freeware, easy to use, platform independent data mining tool. K-means, DBSCAN and various clustering algorithms are already implemented in it. But for this paper, DBSCAN is considered as it efficiently handles irrelevant data.

4. RESEARCH METHODOLOGY
- Take a data set .

- Then, generate test cases using Selenium tool.
- Save test cases that have been generated in an excel file.
- Convert the excel file into attribute related file format(arff) according to the specifications in Weka.
- Load the converted arff file into Weka.
- Apply DBSCAN algorithm to the above loaded data .
- Save the clustered test cases in arff file.
- Take the clustered arff file and again load it into Weka to show the cluster assignments.
- Then, apply appropriate filter to eliminate irrelevant test cases.
- Save output to an arff file.

5. EXPERIMENTAL RESULTS

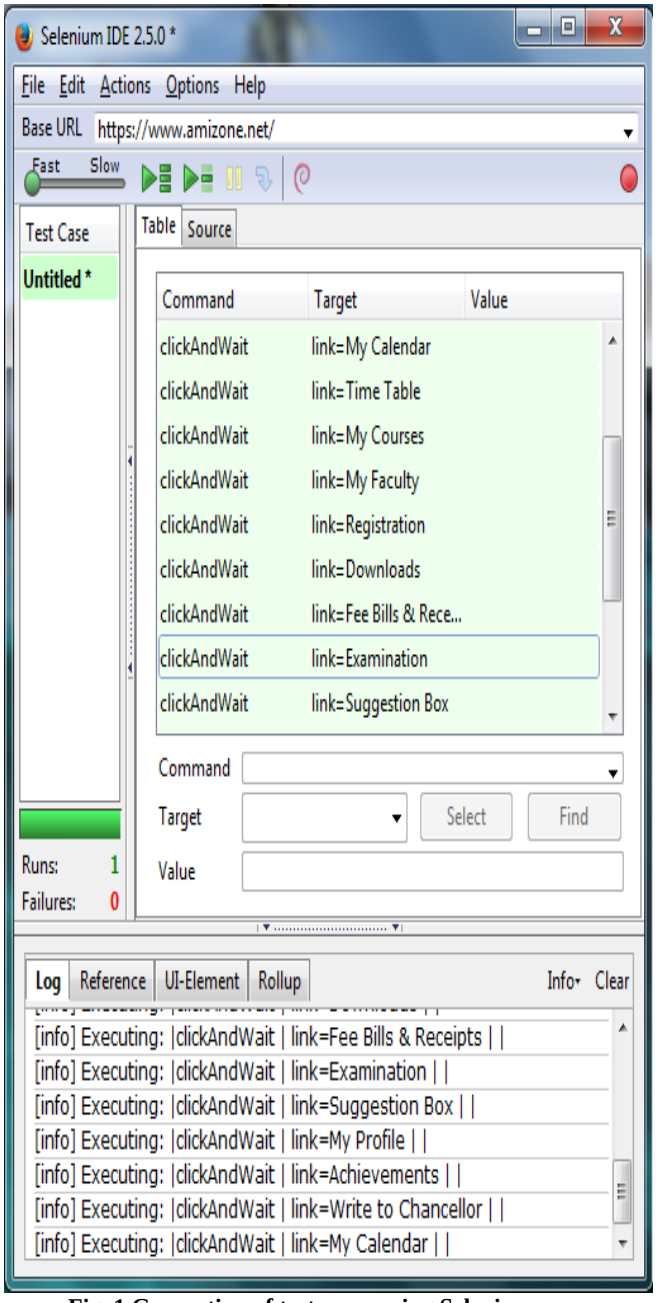

Fig .1 Generation of test cases using Selenium

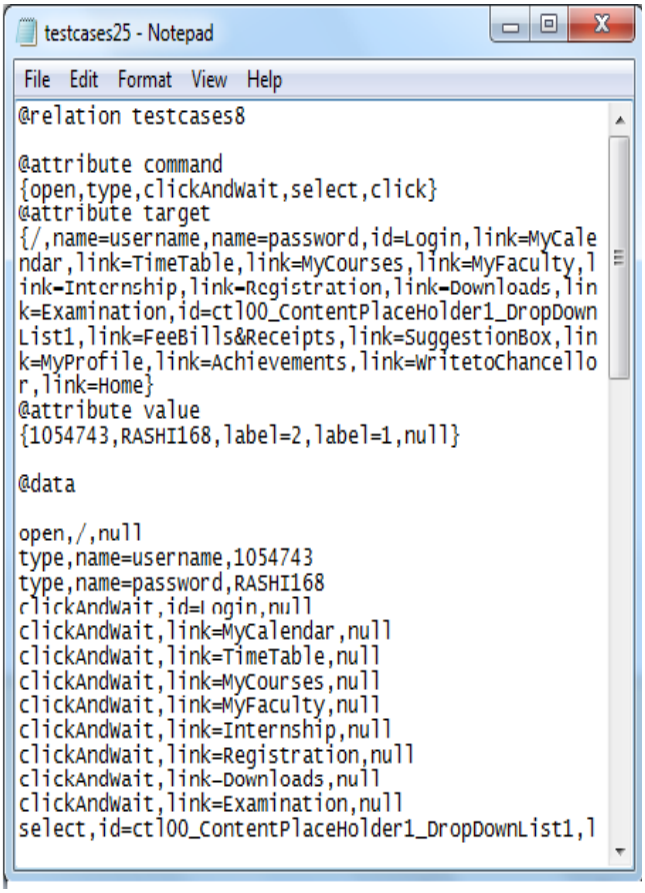

Fig.2 Arff file of test cases

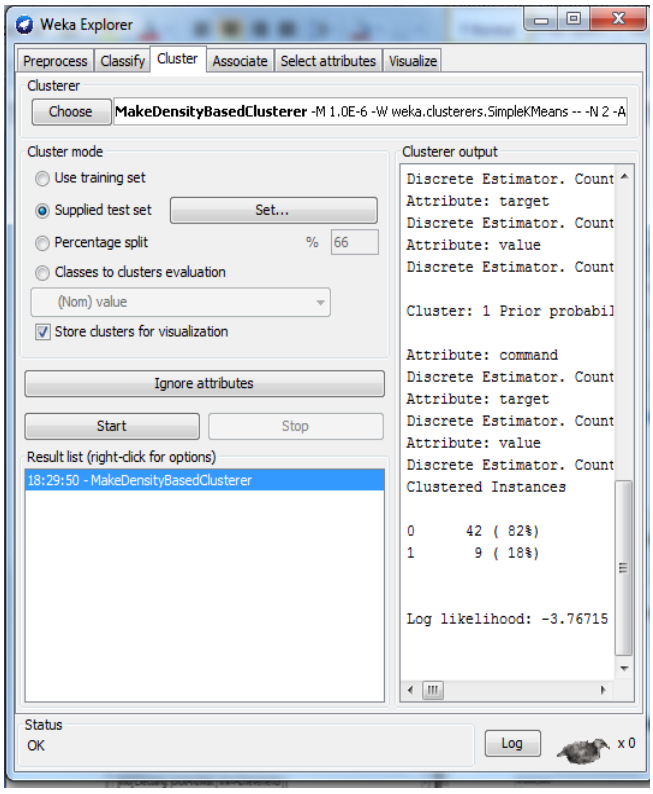

Fig .3 Loading arff file into weka and applying DBSCAN algorithm

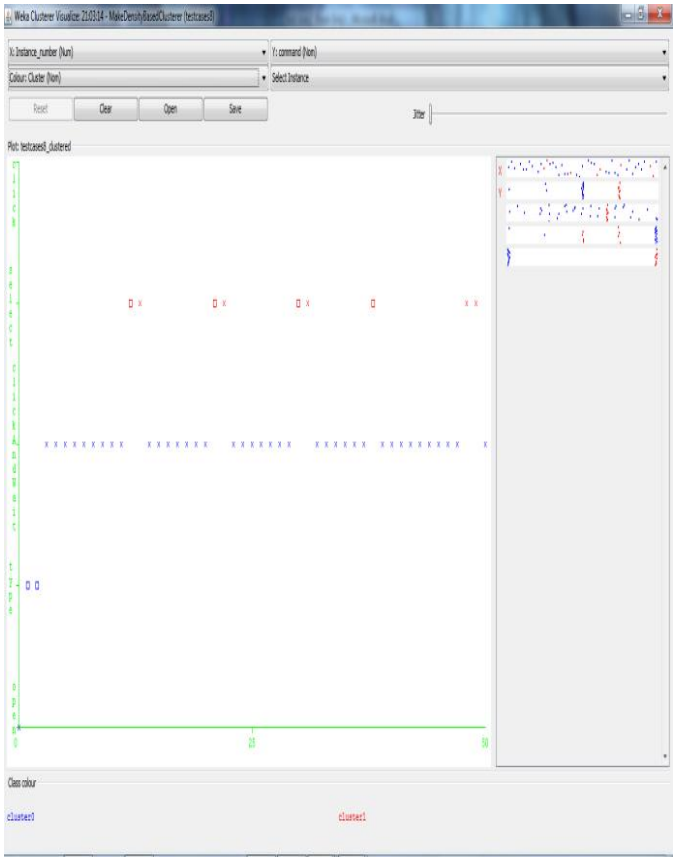

Fig. 4 Plot matrix for visualizing cluster assignments

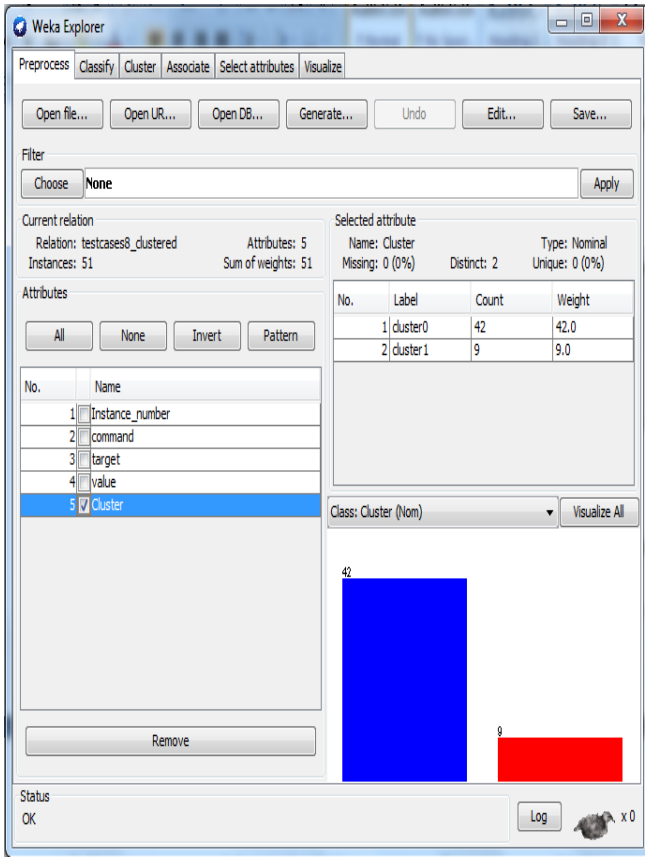

Fig. 5 DBSCAN results

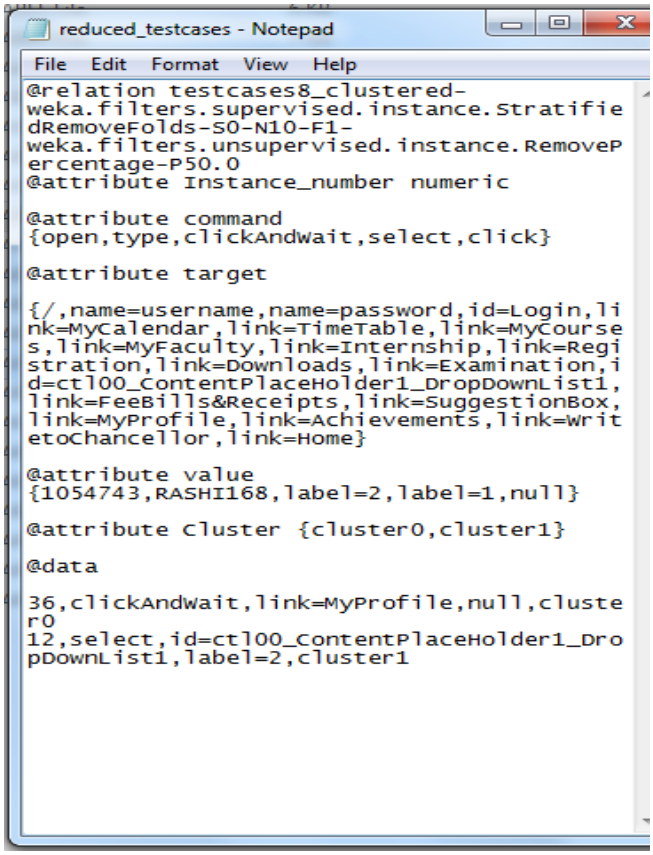

Fig. 6 arff file of reduced test cases after applying appropriate filter

6. PERFORMANCE EVALUATION

Table 1 : Performance evaluation

Clustering algorithm	Number of instances	Test mode	Number of clusters generated	Clustered instances	Time taken to build the model	Incorrectly clustered Instances
DBS-CAN	51	User supplied test set	2	2	0.02 s	6

7. CONCLUSION AND FUTURE WORK

In this paper, a new approach for test suite reduction using density based clustering technique has been discussed. Firstly, generation of test cases from Selenium tool, secondly loading them in Weka and applying DBSCAN clustering algorithm on them and finally removing redundant test cases with the help of appropriate filter. DBSCAN clustering algorithm is not suitable for high-dimensional data .So, our future work deals with generalization of this approach to make it work fine with any test suite.

8. ACKNOWLEDGEMENTS

I would like to acknowledge Ms. Pooja Batra (Lecturer, Computer Science & Engg. , Amity University Haryana, India) and Ms. Sarika Chaudhary (Lecturer, Computer Science & Engg. , Amity University Haryana, India) for their constant help during my research work.

9. REFERENCES

[1] Lilly Rameesh and G.V. Uma, "Reliable Mining of Automatically Generated Test Cases", IJCSI International Journal of Computer Science Issues, Volume 7, Issue 1, January 2010.

[2] Lilly Rameesh and G.V. Uma, "An Efficient Reduction Method For Test Cases", International Journal of Engineering Science and Technology, Volume 2, Issue 11, January 2010.

[3] Kartheek Muthyala et.al., "A Novel Approach To Test Suite Reduction Using Data Mining", Indian Journal of Computer Science and Engineering(IJCSE),Volume2,Issue 3,July 2011.

[4] Ying Huang and Lu Lu, "A Methodology for Test Suit Reduction in User session-based Testing",IEEE,2010.

[5] Yasser El-Sonbaty et.al. ,"An Efficient Density Based Clustering Algorithm for Large Databases",IEEE 2004.

[6] B. Borah and D.K. Bhaftacharyya**, "**A Clustering Technique using Density Difference", IEEE - ICSCN 2007, MIT Campus, Anna University, Chennai, India. Feb. 22-24, 2007. pp.585-588.

[7] Md. Asikur Rahman et.al. , "Density Based Clustering Technique For Efficient Data Mining",IEEE 2008

[8] R. Robu et.al. ,"Arff Convertor Tool for WEKA Data Mining Software",IEEE 2010.

[9] Bin Jiang et.al. ,"Clustering Uncertain Data Based on Probability Distribution Similarity", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013 751.

[10] Dianxiang Xu et.al. ,"Mining Executable Specifications of Web Applications from Selenium IDE Tests**",** 2012 IEEE Sixth International Conference on Software Security and Reliability.

[11] Glory H.Shah, "An Improved DBSCAN, A Density Based Clustering Algorithm with Parameter Selection for High Dimensional Data Sets", 2012 NIRMA UNIVERSITY INTERNATIONAL CONFERENCE ON ENGINEERING, NUiCONE-2012, 06-08DECEMBER, 2012

[12] Chetan Dharni et.al. , "An Improvement of DBSCAN Algorithm to AnalyzeCluster for Large Datasets",IEEE 2013

[13] Remco R. Bouckaert, "Weka Manual 3-6-1",Software manual,June 4,2009,pp-11-14

[14] Jiawei Han, Micheline Kamber, Jian Pei ,"Data Mining Concepts and Techniques", third edition, 2012