ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Efficient Approach of Data Reduction Techniques for Bug Triaging System

Smita Boharpi[1], Prof. Sonal Fatangare[2]

M.E Student, Dept. of Computer Engineering, RMDSSOE Warje, Pune, Maharashtra, India

Professor Dept. of Computer Engineering, RMDSSOE Warje, Pune, Maharashtra, India

ABSTRACT: Bug Triaging is an essential activity of fixing bugs in software development organizations. It is method of allocating a correct developer for fixing a bug. Usually in software development, new bugs are manually triaged by skilled developer. Due to the large number of daily bugs and the lack of expertise of all bugs, the manual bug triage is affluent in time cost and low in accuracy. To decline the expensive cost in manual bug triage, an automatic bug triage approach is used. For bug triage data reduction techniques is used to construct a small scale and high superiority set of bug data by removing bug reports and words which are redundant or not-useful. So we are using instance selection and feature selection concurrently with historical bug data sets. We have added a new module here as feedback session. In this system we will take the normal feedback from the developer after completing the bug.

KEYWORDS: Bug Triage; Data Reduction; Bug Repositories; Prediction for Reduction Order; Feature Selection;Instance Selection; Word Dimension; Bug Dimension.

I. INTRODUCTION

Data mining technology is used in software development process can not only increases the accuracy and completeness of software development but also increases the reliability of the software. A software bugs is an error or fault in a computer program or system that reasons it to create an incorrect or unexpected result. Most bugs come from mistakes and errors made by people in either a program's source code or its design and a few are triggered by compilersproducingincorrectcode. Reportsregarding bugs in a program are generally called as bug reports or fault reports. The main domain of mining bug repositories which has goals to employ data mining to deal with software engineering problems. Software repositories havelarge-scale databases which are used for storing the output of software development. Usually for huge-scale and complex data in software repositories, software analysis is not completely suitable. So data mining techniques, mining software repositories can discover fascinating data in software repositoriesand solve real world software problems [1].

A bug repository is also known as software repository which is used storing information of bugs. The bug triage is essential steps for fixing a bug which are appropriately assigning a developer tonewbug. Foropensourcehuge-scale software projects, the number of day-to-day bugs is so enormous which creates the triaging process very hard and challenging. Software companies pay most of cost in fixing bugs. In a bug repository, a bug is maintained as a bug report, which records the textual description of replicating the bug and updates according to the status of bug fixing. In bug repository, bug reports are called bug data. There are two leading difficulties in software development associated to bug data that may pretend on bug repositories that are the huge scale and the low Superiority. Because of day-to-dayreported bugs, enormous number of new bugs is stored in bug repositories. And low quality bugs are noisy and redundancy. Noisy bugs may worthless data that are related developers and redundant bugs means the same attribute may have different name in different database. They discarded limited time of bug handling [1].

The main goal of data reduction for bug triage to construct a small scale and high superiority set of bug data by removing bug reports and words which are redundant or not-useful. So instance selection and feature selection techniques are used at the same time to reduce the bug dimension and the word dimension. The reduced bug data have small number of bug reports and smaller number of words than original bug data. And they also provide analogous data than original bug data. The instance selection means subset of related instances i.e. bugreport inbug dataand the feature selection means subset of related features i.e. words in bug data [1].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

II. RELATED WORK

In 2011, C.Sun, D.Lo, S.C.Khoo and J.Jiang, [2] they used a bug tracking system. In that system diverse testers or users may submit several reports on the same bugs, denoted as duplicates, which may cost additional maintenance in triaging and fixing bugs. To classify such duplicates or same bugs exactly In this paper propose a retrieval function (REP) to measure the resemblance between two bug reports. It fully employs the data existing in a bug report including not only the resemblance of textual content in summary and description fields, but also resemblance of non-textual fields such as product, component, version, etc. The drawback of that system is there is no indexing structure of bug report repository to speed up the retrieval process.

In 2012, P. S. Bishnu and V. Bhattacherjee [3] purpose software flaw prediction using quad tree-based k-means clustering algorithm. In that paper process the flaw in data using quad treebasedK-meansclusteringtosupportdefect prediction. In that software metrics envisage a value for individual software artifact (e.g. Source code file). The software artifact contains fault according to the extracted features of the artifact.

In 2013, Mamdouh Alenezi and Kenneth Magel, Shadi Banitaan [4] purpose efficient bug triaging using text mining. In this paper, they examine the use of five selection methods forthe correctness of bug assignment. The burden reallocates the load of encumbered developer.

In 2013, S.Shivaji, E.J.Whitehead, Jr. R. Akella, and S.Kim [5] purpose decreasing features to increase code change based bug prediction. In that paper a framework to examine multiple feature selection algorithms and remove noise feature in classification-based defect prediction. It's does not contain how to measure the noise resistance in defect prediction and how to defect noise data.

In 2015,Jifeng Xuan, He Jiang, Yan Hu, Zhilei Ren, Weiqin Zou, Zhongxuan Luo, and Xindong Wu [1] ,in that paper bug triage approach is given. The bug triage is process of correctly assign developer for fixing the fixing a bug. The data reduction techniques are used for high quality data.

III. IMPLEMENTATION DETAILS

A. Problem Definition:

For reducing the affluent cost of manual bug triage we used automatic bug triage method. To build a predictive model for a new bug data sets that present the problem of data reduction for bug triage.

B. Proposed System:

In that system bug data set is input of the system. The bug data set consists of software bug. Each bug has bug reports and the details of the developer. They work on that corresponding bug. The defect report has two parts summary and description. It gives predicted result in the form of output.

In proposed system we are performing data reduction on bug data set which will minimize the scale of data and improve the superiority of data. So that combines instance selection and feature selection to concurrently decrease data scale on the bug dimension as well as word dimension. In that system when a error or a new bug report is encountered we have to assign a developer to fix the bug. Using a classifier indicate to which developer we need to assign the bug.

This can be done by retrieving the information from the history table where we can extract the details like which developer has fixed the bug efficiently, how much time each developer has taken to fix each bug and how many fixers were needed to fix a particular bug etc. Now we shrink the data scale in the bug data sets by using instance selection and feature selection algorithms.

The Feature selection aims to obtain a subset of relevant features by eliminating the uninformative words in the bug reports and the instance selection are used to obtain a subset of relevant instances. Hence by combining both these algorithms we get a reduced bug data set and exchange it with the original bug data. Now this information is sent to the classifier and when a novel bug report is encountered the classifier appropriately allocates developer.

C. System Architecture:

The aim of bug triage is to assign a developer for bug fixing. When a developer is assigned to a new bug report he decides to fix the bug or try to resolve it. The data reduction is used for shrink the scale and increase the superiority of data in bug repositories. The data reduction applied as a phase in data preparation of bug triage. By using instance

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

selection as well as feature selection that reduce data.So that we get high superiority data. For data processingthe instance selection and feature selection are widely used. For given data sets instance selection is to obtain a subset of related instances that is bug report in bug data and feature selection means to obtain a subset of related features that is words in bug data.

Instance selection is procedure to diminish the number of instances by eliminating noisy and redundant instances.Feature selection is a pre-processing method forselectingadiminishsetoffeaturesforhugescale data sets. Merging both these methods we acquire a compact bug data set and exchange it with the novel bug data set. Now this data is conducted to the classifier and when new bug report is encountered the classifier appropriately allocates a developer.
We add new module as feedback session. In this system we will take the normal feedback from the developer after completing the bug. In this feedback session we will take domain, issue, priority, changes, and status. By using this information next time system will select the proper priority and domain for that particular bug. Advantages of that system are selecting the proper domain for the bug; selecting the proper priority for the system; load on the bug system will be added.

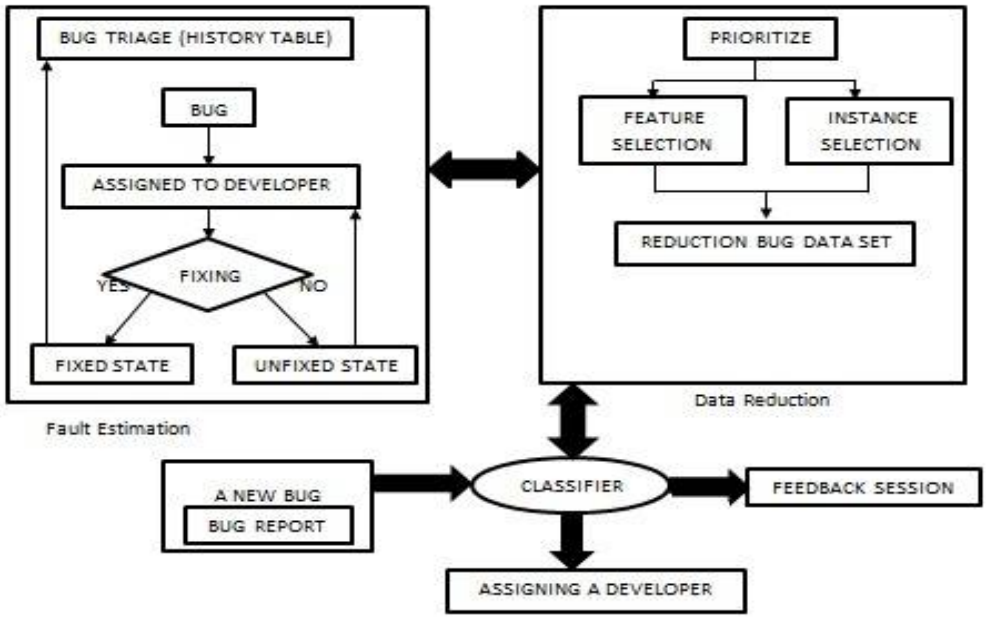

Fig.Architecture of Bug Triage

IV. PROPOSED ALGORITHM

This algorithm is used for data reduction in bug fixing.

Input: T is training set with n words and m bug reports
Reduction Order FS → IS
nF is final number of words
mI is final number of bug reports
Calculate feedback ft for tester

Output: Reduced data set TFI for bug triage

1] Apply FS to n words of T and calculate objective values for all word

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

2] For (inti=1; i<n;i++)
3] {
4] While(ReadFile)
5] Read file and calculate objective values;
6] }
7] Select top match and generating training set TF
8] For (inti=0;i<n;i++)
9] {
10] Find file match word
11] Add into training set
12] }
13] Compare result with bug reported
14] Result Match with Bug Reported
15] If(result<=mI)
16] {
17] Generated final training set and sends to respected developer
18] }

V. RESULTS AND DISCUSSION

Our bug triage system is basically used to track the bug's information occurring in any software. In general any bug is assigned to developer for rectification manually considering the domain of bug, sometimes developer cannot solve so it will be transferred to better developer.

Our system overcome this problem by mining on that bug description, we are processing this bug description removing its stop words, finding the main keywords classifying its domain. After getting domain we assign that bug to the developer who is capable of solving that, we are maintaining developer's quality for assigning. This will be done automatically so manual overhead will be reduced.

A] Practical Work:

Input: Bug dataset and bug report is used as a input dataset in the proposed scheme

1) Login Window- This shows the login form which consist of user id and password and then record is saved successfully.

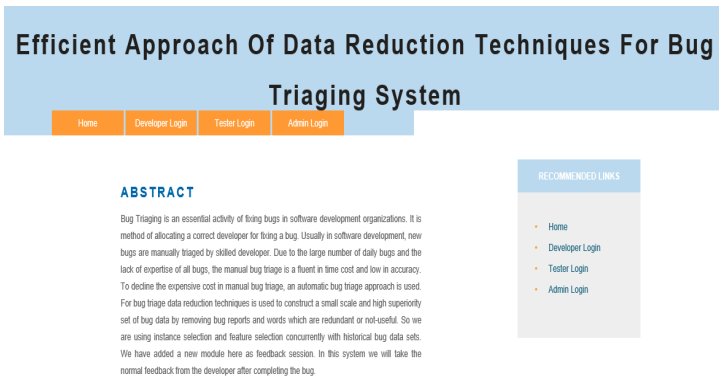

Fig.Home page

2) Upload Dataset- This shows the dataset upload form when clicked on the upload dataset then shows the dataset

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

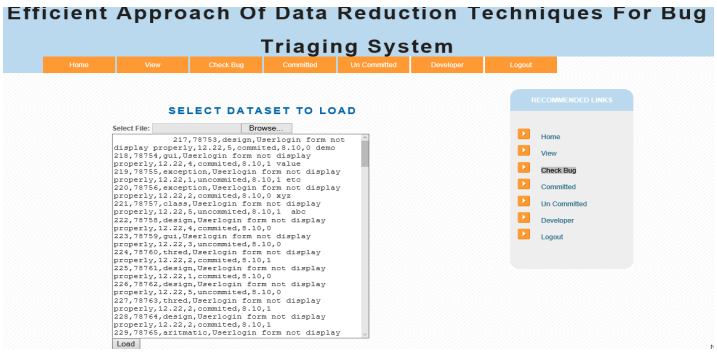

Fig.Upload Datasets

B] Result:

The graph shows the accuracy of the proposed scheme and the existing scheme.It shows that the accuracy of the proposed system is more than the accuracy of the existing system.
Following graph shows the comparison of the proposed system over the existing system.

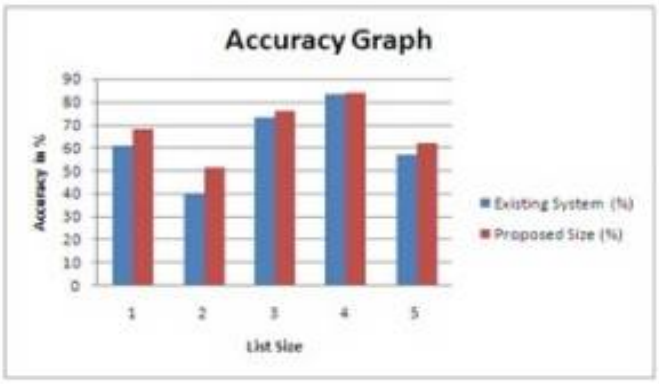

Fig.1: Accuracy Graph

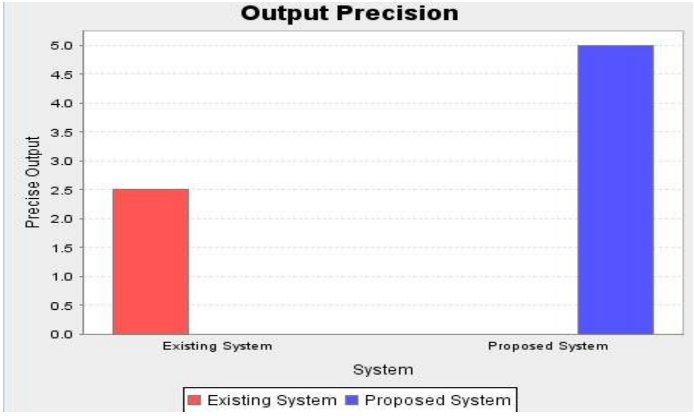

Fig.2.Precision Output of System

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

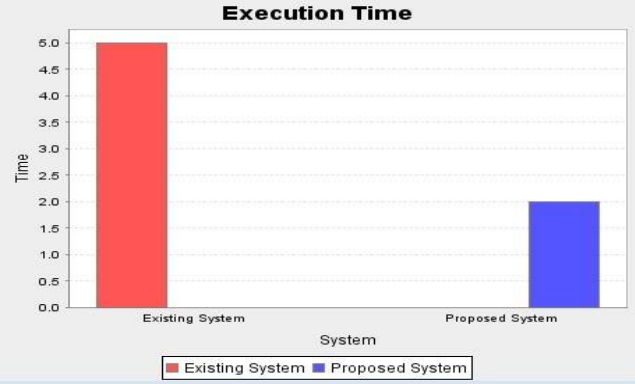

Fig.3 Execution Time

VI. CONCLUSION AND FUTURE WORK

Bug triage is a vital step of software maintenance to save labor cost and time cost. To decrease the expensive cost of manual bug triage we used automatic bug triage approach that appropriately assigns a developer to a new bug for additional usage. The main aim of this work is to reduce the large scale of the training set and to remove the noisy and redundant bug reports for bug triage. In this system engrossed on reducing bug data set in order to have a fewer scale of data and also superiority data. So feature selection and instance selection are used for shrink the scale of bug data sets and also increase the data quality. Using instance selection and feature selection for new bug data set, extract the attributes of each bug data sets and also train a predictive model which is based on historical data sets. For reduced and high quality of data bug data, we used data preprocessing.

In future work we design high superiority bug data sets for bug triage and also increasing result of data reduction.

REFERENCES

1. Jifeng Xuan, He Jiang, Yan Hu, Zhilei Ren, Weiqin Zou, Zhongxuan Luo, and Xindong Wu,"Towards Effective Bug Triage with Software Data Reduction Techniques ,in IEEE transactions on knowledge and data engineering", Vol. 27, No. 1, January 2015.
2. C. Sun, D. Lo, S. C. Khoo, and J. Jiang, "Towards more accurate retrieval of duplicate bug reports", in Proc. 26th IEEE/ACM Int. Conf. Automated Softw. Engg., 2011, pp. 253262.
3. P. S. Bishnu and V. Bhattacherjee, "Software fault prediction using quad tree-basedk-means clustering algorithm", in IEEE Trans.Knowl. Data Engg., vol. 24, no. 6, pp. 11461150, Jun. 2012.
4. Mamdouh Alenezi and Kenneth Magel, Shadi Banitaan" Efficient Bug Triaging Using Text Mining", 2013 academy publisher data sets.
5. S. Shivaji, E. J. Whitehead, Jr., R. Akella, and S. Kim, "Reducing features to improve code change based bug prediction", in IEEE Trans. Soft. Engg., vol. 39, no. 4, pp. 552569, Apr. 2013.
6. G.Parthasarathy, D.C.Tomar, Blessy John"Analysis of Bug Triage using Data Preprocessing(Reduction) Techniques", in International Journal of Computer Application(0975-8887) Volume 125-No.9,Sept.2015.
7. Weiqin Zou,Xin Xia,Weiqiang,Zhenyu Chen,and David Lo ,"An Empirical Study of Bug Fixing Rate", in Department of information Engineering, Jiangxi University of Science and technology,China.
8. WeiqinZou,XinXia,Weiqiang,ZhenyuChen,and David Lo ,"Towards Training Set of Reduction For Bug Triage", in Department of information Engineering, Jiangxi University of Science and technology,China.

BIOGRAPHY

Smita Boharpi received B.E. degree in Computer Science and Engineering in 2008 from Dr.Babasaheb Ambedkar Technological University, Lonere, Dist.Raigad andpursuing M.E.from RMDSSOE, Warje, Pune**.**

Sonal Fatangare is working with RMDSSOE, Warje, Pune as an Assistant Professor. She has experience of 5 yrs. in the field of teaching and research and her research interests are Network Security and Data Mining.