

PRIORITIES OF DEVELOPERS BASED ON INSTANCE SELECTION AND FEATURE SELECTION TECHNIQUE

Ms.Anchal R.Ghutake¹

Prof. Amreen Khan²

1 M-Tech Scholar, Department of computer science and engineering, G.H.R.C.E, Nagpur, India, anchalghutake17@gmail.com

2 Professor, Department of computer science and engineering, G.H.R.C.E, Nagpur, India, amreenkhan@raisoni.net

Abstract - To deal with Large number of software Bugs every day in software companies. The process of Bug fixing used the important step is Bug triage. Assign a Bug report to the correct potential developer is the main moto of Bug triage. The previous system using machine learning algorithms ,which is classification of text bugs from the sets of bug reports. In one of the most important software repositories, developers update and create bug reports to corroborate software maintenance. In this paper,we find the fix developer, which aims to improve rank of developers. we collect data set for bugs of large open source projects such as eclipse and calculating fix developer of particular bug using feature selection technique. To resolve the directive of using instance selection technique for classification bug data set. Our system improve the accuracy of bug triage and reduce the bug dimension such as redundancy find out of total Bug data set.

Keywords - feature selection, instance selection, Bug data reduction, fault localization

1. Introduction

Now a day's all IT companies are influenced by software bugs. In India, Many software and IT companies waste most of the money in the wastage of Bugs. To Collect all the information related to bugs containing bug repository in the software projects for storing output eg.bug, source code, emails .bug fixing is time consuming and significant process in software maintenance. Increasing the large number of day to day bugs therefore it is so impossible to handle them without gap. software bugs are affected in all software projects. The pure quality and large scale of bug data are two the defiance in software development. Some software maintenance show that the costs is more than 80%,of the original costs include with software product. we present the different types of problem of data reduction for Bug triage Everyday reported the new and different type of bugs are appear in software projects. The main approach of our system provides solution to open source tool like eclipse for managing the bug data such redundancy, time cost and efficiency .Many open source tool like Mozilla, GNOME, Firefox these are the applications containing hundreds of bug reports a per

day, basically each bug assigned to the developer who can fix it at minimum amount of time is a main Moto in our system. This process of assigning bugs to the proper developer, known as Bug triage. The proposed system is designed using a feature selection and instance selection technique to find fix developer to each of bug data set. In eclipse project ,many bugs are assigning to To some feature selection technique using x2-test (CHI) algorithm are developer to assigned automatic Bug triage.

2. RELATED WORK

From the study on various types of techniques providers some of them are illustrated as, The reduced bug data contain bug reports and words than provide similar information over the original bug data. Some conventional methods of accuracy of bug data set using in different type applications. Literature survey includes various techniques and work done related to bug data set

SFifing Xuan have presented text classification techniques in which SVM algorithm are used to reduce the dimension of bug data sets as well as improve the data quality and also this technique classify the bug data set to get the fix developer of particular type bug. It can enhance the accuracy by removing bug reports and calculating the 50% efficiency of bug data set from original bug data set of Firefox tool.[1] Shay Artzi [2] The proposed system having implementing the Apollo's algorithms, for generating new type of bug data set automatically classify into the fix developer by using PHP programming language. In order to get survey on MOZILLA,APPECHE and ECLIPSE developers. This type of tool can 41% of Bug report solve by developer in 2014.

Peter zon[3] proposed the tabu search condensing algorithm to remove the space of all sets of bug reports and stress the size decrease and accuracy of bug data set. .Dominique Matter [5] proposed work based on collecting the program code and using fault localization technique then vocabulary found in the difference of a programmers collecting the characterization of a bug report. Shivkumar Shiva [6] have proposed the feature selection technique use to improve faster bug predictions and extract the feature from previous bug dataset.Classifier-based bug prediction techniques such as feature selection method used to extract relevant and

irrelevant features of bug data set should be allowed fast, precise, and accurate bug predictions.

Thomas Zimmermann[7] have proposed the work on pre-processing of bug dataset by using a combining the methods such as machine learning and text classification method checked and confirm for bug reopens based on 350 Microsoft developers and improve the analysis of bug reports. this two methods improve the comparative analysis of bug dataset. Gail C [8] This technique apply into the Mozilla application then sort out the raw of bug data and the pure quality of bug data is reduced. In this approach on bug reports, achieving 40% efficiency with current type of bug set.

[9] Basically analyze the two challenges, such as modeling the developer ranking in a bug repository and assigning the bug to the fix developer. In this approach, there is accuracy of bug data set from GNOME tool. [10] R. Abreu proposed and defined An automated technique such as Program-state based fault localization method using a novel analysis of the history of the source code determine how many bugs are resolve and unresolved by the developer. During preprocessing of bug-data ,developer spending a high amount of effort of Bug dataset then check into thoroughly the bug in the order to give answer of questions about it

3. III Methodology

3.1 Block Diagram:

The main idea behind this chapter is to give brief idea about designing the such type of system having fix developer to particular bug data set and update the calculate ranking of each developer in bug data set of the open source tool like ellipse. Such type of system to know the proper developer of assigning the bug data. The overall Design approach is basically divided into three Levels. In three levels process first defines the preprocessing process in Bug data set of eclipse. This is the main phase of converting the raw of bug dataset into a useful of bug data set. In this process the redundancy is calculating by the data set in which the repeated type of bug eliminating into the bug data set.

In the second level defines the classification of Bug data set. For doing the classification we used instance selection technique like iterative case filter (ICF) algorithm in which we do the classification of bug data set on the basis of eclipse browser. In this process the classification of bug dataset into product and company wise then show the data set classify into grouping of product and company of bug dataset. further in the third and last level of bug data set to calculating

the ranking of developer on basis of feature selection technique like x2-test algorithm(CHI) algorithm

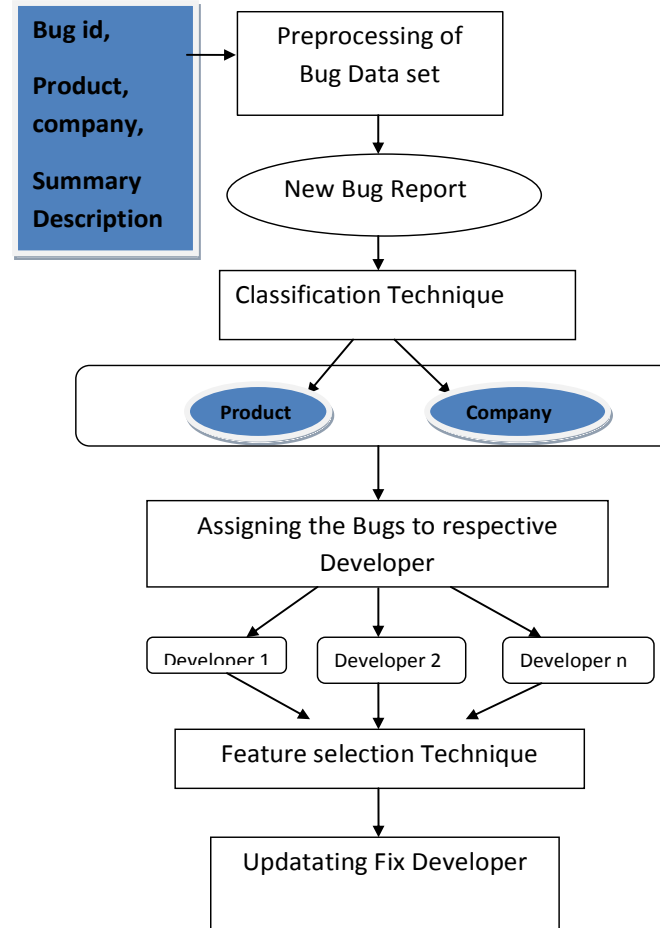


Fig 3.1.1: Block Diagram of Bug triage

A. Preprocessing of bug data set:

In preprocessing module we are going to create the useful bug data set and finding the redundancy by using fault localization technique. Calculate preprocessing time and final bug data dimension for each bug data . In the preprocessing module, If the bug data is repeated it will take only once and calculate the redundancy as follows:

$$\text{Redundancy} = \text{Reduce the Bug} / \text{Total bug data set} * 100$$

Initial dimension	Final dimension	Dimension reduction
976	854	12%

Fig 3.1.2 comparison table for initial and final dimension

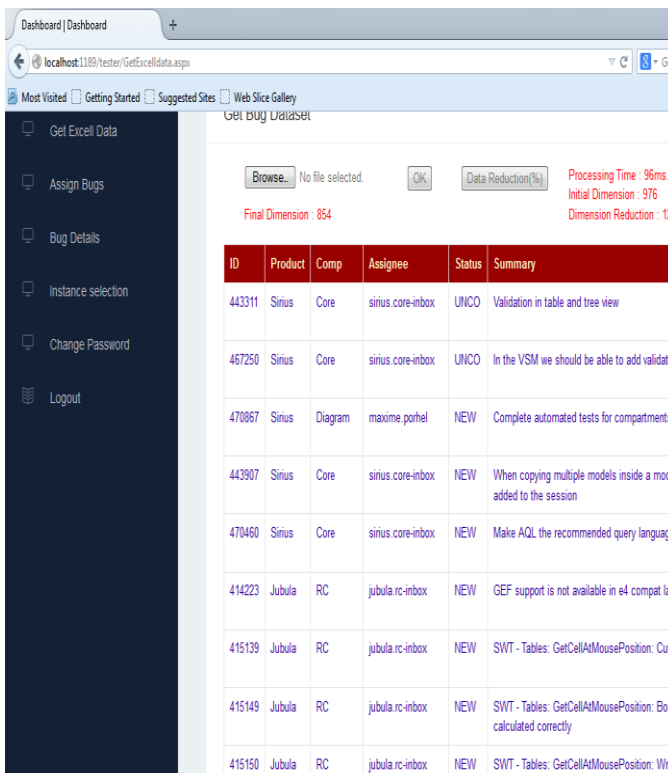


Fig :Data reduction

B. Instance selection technique:

In this project, experimentally shows that the instance selection algorithm is instance classification generate. the most advantage of this algorithm is faster classification accuracy provide and also the represent in minimum space. We are going to use ICF is an instance selection algorithm in which used to evaluate whether an instance is representative for the classes of company and product.

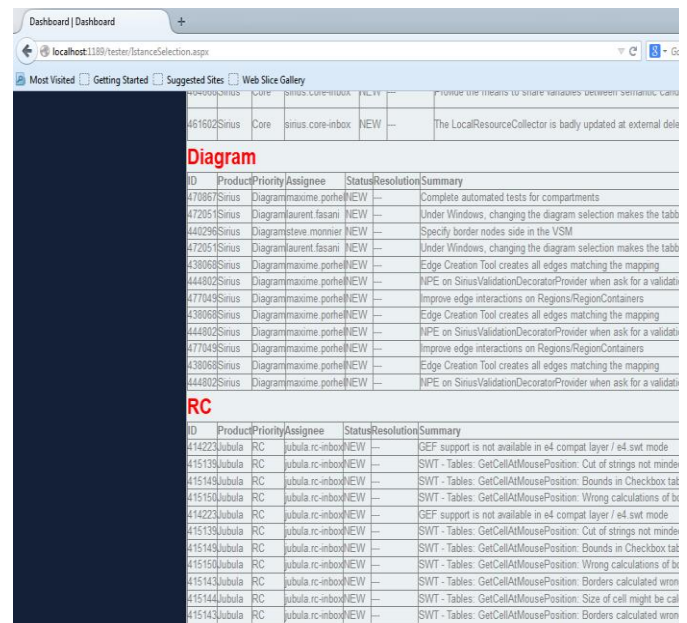


Fig 3.1.3 Classification of Bug dataset

C. Feature selection algorithm:

We are going to create the assigning bug data set to developer then using feature selection algorithm, then this algorithm extract features then finding the how many bug are solve or unsolved. Analyze the process of how many bugs solve by one developer and updating list of the rank of developer. Also calculating the total efficiency of Bug data set.

at that time by the developer then bug are resolve. The result of Bug triage shown in fig, how many resolve bug as well as unresolved bug should be generated

B. Individual Graph:

To calculate the how many bugs are resolve to individual developer for generating the individual graph of each developer as show in fig . then easy to know the which developer is fix.

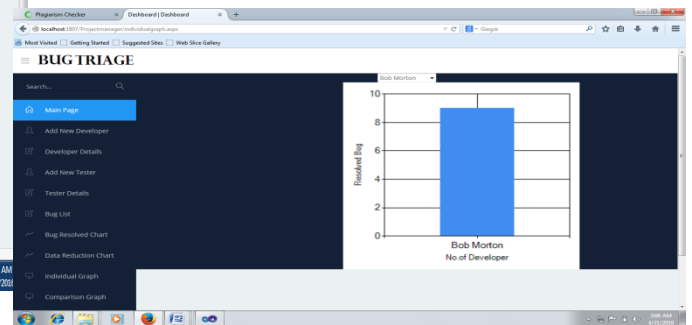
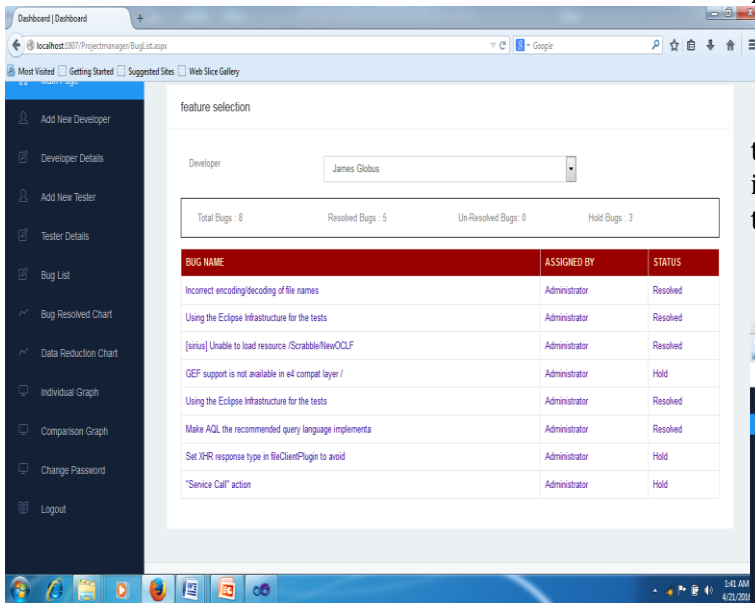


Fig 3.1.4: Total no of resolve and unresolved bug

Fig 4.1.2 Individual graph

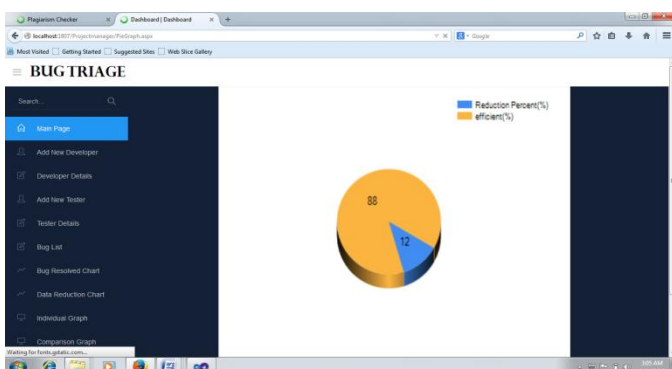
4.Results

A. Data reduction:

In our work to design firstly ,authentication of tester, developer and project manager. registration and authentication of the user and data owner will take place by which the Administrator will provide the user and data owner with an user-id and password. In second preprocessing module, If the bug data is repeated it will take only once and calculate the redundancy. Then fig shows that 12% data reduce into the existing of Bug data set.

C. Bug triage :

Firstly we assign the We are going to create the assigning bug data set to developer then using feature selection algorithm, then this algorithm extract features then finding the how many bug are solve or unsolved. Analyze the process of how many bugs solve by one developer and updating list of the rank of developer .



Sr.no	Name of Developers	Total Assign Bug	No of Resolve Bugs	No of Hold Bugs
1	Guest	14	6	8
2	James Globus	14	5	3
3	Bob	14	9	4
4	Amol	14	5	7

Fig 4.1.1 data reduction

Table 4.1.3 Calculating resolve Bug

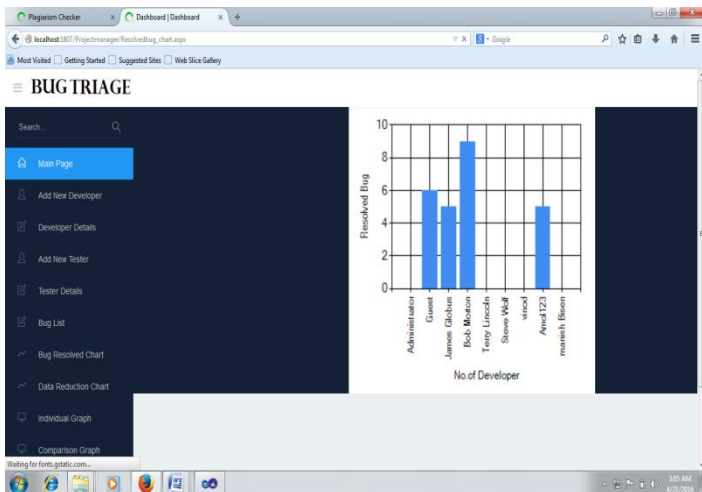


Fig 4.1.4 Fix developer Graph

D. Comparative analysis:

The below figure shows comparative analysis of overall system performance efficiency of the dug dataset. In our work to design and implementing the fix developer to assigned a bug. In existing system work on many algorithm not more than 60% efficiency of Bug data in open source tool like eclipse and we are implemented the ICF and x2-test algorithm based system have 80% efficiency on the comparison on the basis of that.

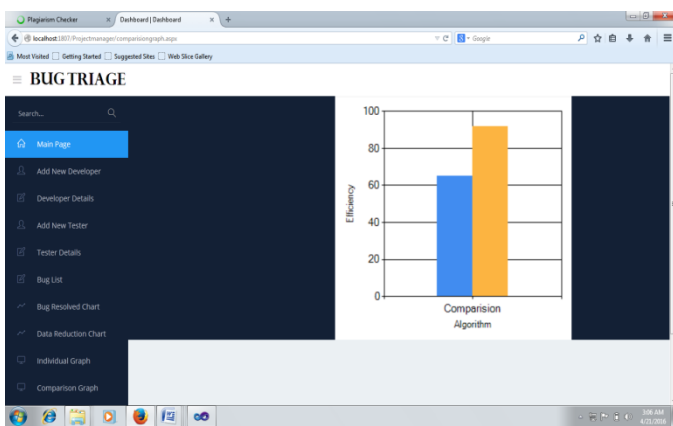


Fig 5.1.1 Comparative Analysis of Existing and Previous System.

So, finally by comparing all these we got the favorable performance of proposed system than the previous system which can be derived from the above graph.

VI. Conclusion

We Proposed the mechanism that is simple to implement and designed the fix developer to assigning a Bug data and reducing the pure , unwanted bug into the dataset. we have executed both the scenarios as classification as well as the feature extract with instance selection algorithm and feature selection algorithm based system. The results are compared with existing system so our proposed system improve the accuracy ,redundancy and maximize the efficiency.

References

1. Vicente cerveron and francesc j. ferri “Another move toward the minimum consistent subset: A Tabu Search Approach to the Condensed Nearest Neighbor Rule”. IEEE transactions on knowledge and data engineering , Vol. 22, No. 1, Feb. 2012.
2. Dominique matter, adrian kuhn, oscar nierstrasz” Assigning Bug Reports using Vocabulary -Based Expertise Model of Developers” IEEE transactions on knowledge and data engineering , Vol. 22, No. 1, Nov. 2011.
3. Ahmed lamkanfi, serge demeyer, emanuel giger , bart goethals” Predicting the Severity of a Reported Bug” IEEE transactions on knowledge and data engineering , IEEE transactions on knowledge and data engineering , Vol.12, No.4, Nov.2010.
4. Shivkumar shivaji ,E. James Whitehead, jr. ram akella ,sunghun kim” Reducing Features to Improve Code Change Based Bug Prediction” IEEE trans, cybern. Vol.31, No 5, May.2009.
5. Thomas zimmermann ,nachiappan nagappan , brendan murphy”characterizing and predicting which bugs get reopened” ieee transactions on knowledge and data engineering , ieee transactions on knowledge and data engineering , vol.15, no.6, aug.2008.
6. R. e. walpole, r. h. myers, s. l. myers, and k. ye, x. wang, l. zhang, t. xie, j. anvik, and j. sun, “An Approach to Detecting Duplicate Bug Reports using Natural Language and Execution Information,” in proc. 30th int. conf. softw. eng., May 2007.
7. M. rogati and y. yang, “High-Performing Feature Selection for Text Classification,” in proc. 11th int. conf. inform. knowl. manag., IEEE transactions on knowledge and data engineering , Vol.12, No.4, Nov. 2007. [10] Murphy-hill, t. Zimmermann, C. bird, anNagappan, “The Design of Bug Fixes,” IEEE transactions on knowledge and data engineering , Vol.12, No.4, June.2007
8. [8] T. m. khoshgoftaar, k. gao, and n. seliya, “Attribute Selection and Imbalanced data:

Problems In Software Defect Prediction,"IEEE transactions on knowledge and data engineering ,Vol.1,No.4, May 2006.

BIOGRAPHIES



Anchal Ghutake pursuing M.Tech. Degree from G.H.R.C, India and also received the B.E. Degree from SGB Amravati University, India and, research interest focused on Core compute science and engg.

Prof. Ms. Amreen Khan
compute science and engg
GHRCE,India
amreenkhan@raisoni.net