# Novel Bug Triage repositories on instance selection

**Maddela Sai Kiran[1], Gogineni Jyothi[2] & Mr. B. Laxmaiah[3]**

[3]M-Tech, Dept of CSE,Sharada Institute Of  Science and Technology, Khammam.

[2]Associate Professor, Dept of CSE,Sharada Institute Of  Science and Technology, Khammam.

[1]HOD & Associate professor Dept of CSE, Sharada Institute Of Science and Technology, Khammam.

## Abstract

Software companies are dealing with large number of software bugs nowadays. It is well expensive and unavoidable too. The traiging process is nothing but to assign effective and proper developer for bug fixing. various techniques are used for this process earlier. Bug triaging and historical data set are well maintained. New methods are being used each time. Human traiging was not a time worth process and was not successful in the case of proper assignment of developers .Nowadays the techniquespay attention towards correct assignment of the developer to the respective bug automatically. It will be an effective method for the companies who face the challenge of assigning the developer.

**Keywords:** Bug, Bug Triage, repositories, instance selection.

## 1. Introduction

Many software companies spend almost half of their project money in fixing the bugs. Large software projects have bug repository that holds all the information related to bugs and is well maintained for further processing. In bug repository, each software bug has a bug report and is also known by bug data. The bug report consists of textual information of the bug and the updates on the basis of the status of bug fixing, which is available in historical bug data set. Traditional software analysis is not fully suitable for the largescale and complex data in software repositories. Data mining has been introduced to the technically developing environment as a promising means to handle the software data. By using the data mining techniques, mining software repositories can uncover interesting and hidden information of the software repositories and can also solve the real world software problems. Due to the huge amount of daily reported bugs, the bug reports are increasing and the scaling up in the repository is being high as well. Noisy bugs and redundant bugs are degrading the quality of bug reports which is held in the repositories. Bug triage is one of the least time taken procedures in handling of bugs in software projects and of course the most proper way. .Manual bug triaging by a human bug triager is a vast process

and error zone because of the arrival of large number of bug data and lack of developers who has an accurate knowledge of the bug to be fixed with. In some former methods, if a bug report is formed or a bug occurred, then a human triager assigns this bug to a developer, who tries to fix this bug. This developer is recorded in an item assigned to in historical bug dataset. If the previously assigned developer was unable fix this bug then the former will change to a new one. The method of assigning a proper developer for fixing the bug is known as bug triaging.

## 2. Related Work

As our Knowledge, there is no combination of data reduction methods in turn to decrease the data scale and upgrade the exactness of bug triage approach in the illustration. Jeong, Kim, Zimmermann introduced a tossing graph model based on Markov property from the conception of reassign the bug reports to other developers [6]. Shivaji and colleagues [12] proposed the feature selection techniques to predict the software bugs. Anvik, L. Hiew, and G. C. Murphy [1] extend the machine learning approaches. They describe the bug triage as semi-supervised approach which updated with weighted recommendation list; based on the probabilistic view the relevant developers are employed to the human triage [4,

13]. Cubranic and Murphy [3] projected supervised learning method (NB Classifier) to assist in bug triage by using text categorization to predict the relevant developers. A classification model should be designed to investigate the relationship among the datas in bug data set and to check the quality [12, 14]. Fu.Y, Zhu.X, and Li.B [4] investigated to obtain the accurate prediction model with minimum cost by labelling most informative instances. In contrast to these papers, our paper aims to employ the information gain algorithm to develop the software value of bug data prediction. In this paper, we focus on the issue of bug data reduction and low in precision of bug data set. Further the combination of feature selection and instance selection algorithm intend to shrink the bug data set and develop the performance of bug triage with high-quality bug data in software maintenance and improvement.

## 3. Implementation

There are two approaches in mining the bug reports in the bug storage. Feature selection method is used first to reduce words and Instance selection is used to remove repeated and blank reports. Thirdly, PSO approach is used to find the expert for fixing the bug

---

**Algorithm 1.** Data reduction based on FS → IS

| | |
|---|---|
| Input: | training set $T$ with $n$ words and $m$ bug reports, reduction order FS→IS<br>final number $n_F$ of words,<br>final number $m_I$ of bug reports, |
| Output: | reduced data set $T_{FI}$ for bug triage |

1) apply FS to $n$ words of $T$ and calculate objective values for all the words;
2) select the top $n_F$ words of $T$ and generate a training set $T_F$;
3) apply IS to $m_I$ bug reports of $T_F$;
4) terminate IS when the number of bug reports is equal to or less than $m_I$ and generate the final training set $T_{FI}$.

---

The algorithm depicts the overall data reduction method in it. Existing bug reports in the bug storage treated as training set and the new coming bugs are the test set

### 1. FEATURE SELECTION

It selects a minimum set of features such that the prospect distribution of different classes given the values for those features is as close as possible to the original distribution given the values of all features[4][5]. Reduce specified number of patterns in the patterns, easier to understand. Select new attributes in the data set that can detain the key information in a data set much more proficiently than the unique attributes. Use the smallest depiction which is enough to solve the task [3].

### 2. INSTANCE SELECTION

Instance selection method is associated with the classification and clustering methods in data mining. It is not easy to find the valid one

without the patterns in the bug report. The attributes of the data are taken and it is classified according to their patterns. It is independent and takes the original data set attributes and finds the reduced data set. Classifier is used to classify the attributes in every bug.

### 3. PARTICLE SWARM OPTIMIZER

PSO is initialized with the group of solutions and then searches for best possible by updating solutions. Particles move through the solution space and select the fitness criterion from each iteration by following the best values it has achieved so far. With that solution, the bug report is fixed by the expert.

### 3.1 MODULE DESCRIPTION

This module show's four part's as follow:

1. Firstly it will show how many bugs are not assigned to any developer. It will give complete status about the bugs to the admin so that he will come to know which bugs are not assigned yet.

2. Secondly it will show how many bugs are not assigned to any developer. It will give complete status about the bugs to the reporter so that he will come to know which bugs are assigned.

3. Thirdly it will show how many bugs are rectified by the developer's. It will give complete status about the bugs to the admin so

that he will come to know which bugs are rectified completely.

4. Fourthly it results the number of bugs are not fixed by the developer's. It will give complete status about the bugs to the admin so that he will come to know which bugs are not rectified yet. Historical data also used for reducing the bug data. Here we selected the recent accessed bugs and the data which we get we will use it for data reduction.
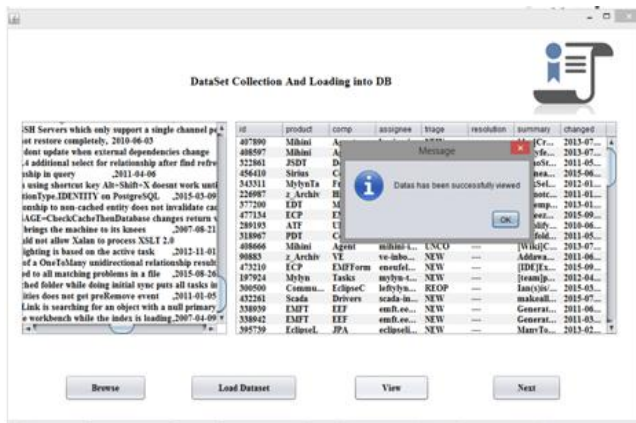
## 4. Experimental Work
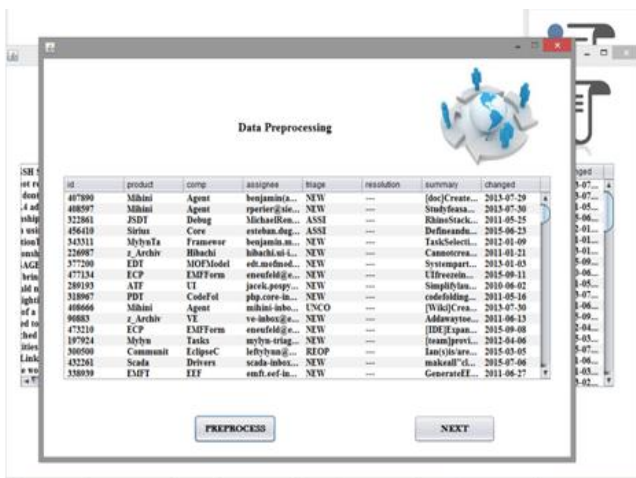


Fig 1: Dataset Collection Loading into DB.
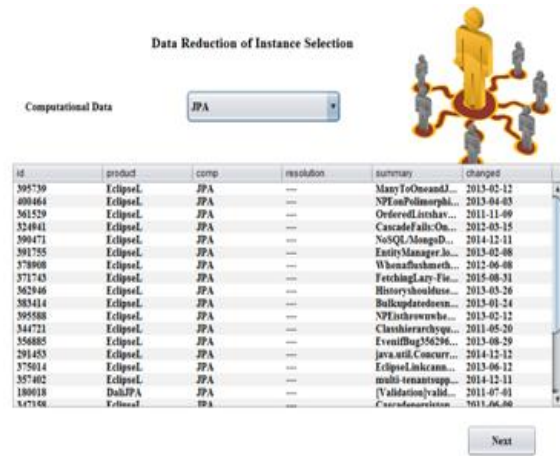


Fig 2: Data Processing.



Fig 3: Data reduction of instance selection.

## 5. Conclusion

One of the expensive steps in software maintenance is Bug Triaging, ma in ly when it comes to the matter of labor and time cost. The recent technique aims to form reduced and high quality bug data in software development and thereby maintenance. The data processing techniques like instance selection and feature selection are used for data reduction. Former techniques were tossing graphs, collaborative approach and semi-supervised learning. The latter system is useful for any open source projects that generate huge and large amount of bug data. Several software companies' doe's projects like banking, food chain management can consider the application of the current technique which is more beneficial and effective.

## 6. References

[1] S. Artzi, A. Kie_ zun, J. Dolby, F. Tip, D. Dig, A.Paradkar, and M. D. Ernst, "Finding bugs in web applications using dynamic test generation and explicit-state model checking," IEEE Softw., vol. 36,no. 4, pp. 474–494, Jul./Aug 2010.

[2] Olga Baysal, Reid Holmes, and Michael W. Godfrey David R. Cheriton ,"Revisiting Bug Triage and Resolution Practices", School of Computer Science University of Waterloo Waterloo, ON,july 2009.

[3] G. Jeong, S. Kim, and T. Zimmermann, "Improving training set reduction for bug triage," in Proc. 35th bug triage with tossing graphs," in Proc. Joint Meeting 12th Eur. Softw. Eng. Conf. 17th ACM SIGSOFT Symp. Found. Softw. Eng., Aug. 2009, pp. 111–120.

[4] J. W. Park, M. W. Lee, J. Kim, S. W. Hwang, and S. Kim,"Costriage: A cost-aware triage algorithm for bug reporting systems," in Proc. 25th Conf. Artif. Intell., Aug. 2011, pp. 139–144.

[5] Katja Kevic, Sebastian C. Muller, Thomas Fritz, and Harald C. Gall "Collaborative Bug Triaging using Textual Similarities and Change Set Analysis", Department of Informatics University of Zurich, Switzerland ,june 2010 katja.kevic@uzh.ch {smueller, fritz, gall}@ifi.uzh.ch.

[6] J. Xuan, H. Jiang, Z. Ren, J. Yan, and Z. Luo, "Automatic bug triage using semi-supervised text classification," in Proc. 22nd Int. Conf. Softw. Eng. Knowl. Eng., Jul. 2010, pp. 209– 214.

[7] D.Cubrani_c and G. C. Murphy, "Automatic bug triage using text categorization," in Proc. 16th Int. Conf. Softw. Eng. Knowl. Eng.,Jun. 2004, pp. 92– 97.

[8] Towards Effective Bug Traige With Software Data Reduction Techniques,Jifeng Xuan, He Jiang, Member, IEEE, Yan Hu, Zhilei Ren, Weiqin Zou,Zhongxuan Luo, and Xindong Wu, Fellow, IEEE,IEEE transaction on Dta Knowledge and Engineering.

[9] 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 2008,pp. 461–470.

[10] J. Xuan, H. Jiang, Z. Ren, and Z. Luo, "Solving the large scale next release problem with a backbone based multilevel algorithm,"IEEE Trans. Softw. Eng., vol. 38, no. 5, pp. 1195–1212, Sept./Oct.2012.

[11] J. Xuan, H. Jiang, Z. Ren, and W. Zou, "Developer prioritization in bug repositories," in Proc. 34th Int. Conf. Softw. Eng., 2012, pp. 25– 35.

[12] Y. Yang and J. Pedersen, "A comparative study on feature selection in text categorization," in Proc. Int. Conf. Mach. Learn., 1997, pp. 412–420.

[13] H. Zhang, L. Gong, and S. Versteeg, "Predicting bug-Fixing time:An empirical study

of commercial software projects," in Proc. 35th Int. Conf. Software Eng.,May 2013, pp. 1042–1051.

[14] W. Zou, Y. Hu, J. Xuan, and H. Jiang, "Towards Annu. IEEE Int. Comput. Soft. Appl. Conf., Jul. 2011, pp. 576–581.

## Authors Profile

**MADDELA SAI KIRAN**

B-Tech in Swarna Bharathi Institute Of Science And Technology, Khammam

M-Tech[CSE] (Computer Science And Engineering)Sharada Institute Of Science and Technology, Khammam.

Mail id: kiran.sai.maddela@gmail.com

**GOGINENI JYOTHI.**

Experience: 10 years.

Qualification: M.Tech from JNTU, Hyderabad.

Designation: Associate Professor.

Working: Sarada Institute of Technology & Science(SITS), Khammam.

Emaild: jgogineni@gmail.com

## HOD Details

**Mr. B. Laxmaiah**

Head of the Department, Associate professor CSE, Sarada Institute of Technology & Scinece (SITS), Khammam. He obtained M.Tech degree from JNTUH, Hyderabad. His research areas include Object Oriented Programming Througt Java, Data base Management System, Data Structures, Web Services, Data Warehousing and Data Mining and Operating Systems.