

A Survey on Software Bug Triage with Bug Reduction Mechanism

Thakare M. R.¹, Ahuja M.M.², Jagdale S.B.³, Asst. Prof. Vasava V.M.⁴

^{1, 2, 3, 4} Department of Computer Engineering

^{1, 2, 3, 4} NES GCoE's, Nagaon, Dhule (MS) -424005

Abstract- Now days, every software industry are suffering lot of cost and efforts empowering the software bugs and software developer try to make solution for fixing the bug is a bug triage. Which is aims to specifically assign the new bug to developers. It helps decrease time and cost in the field of manual work. We empirically investigate the performance of data reduction on totally 600,000 bug reports of two large open source projects, namely Eclipse and Mozilla. In this research paper, we focused on the problems on data reduction for bug triage using reliable technique to handle automatic bug triage is text classification technique. Also prepare that how to reduced the scale of data and improve the quality of bug data. Using combination of instance selection and features selection to equally reduced the data scale on bug dimension and word dimension. To determine these both selections we extract attributes from historical bug data sets and build a predictive model for a new bug data set.

Keywords- Bug Triage, Text Classification, Bug Dimensions, Word Dimensions.

I. INTRODUCTION

Now a day's lot of software industry more over finance is use to fixing bug. Most of the bugs are save the all contents in bug repository. So, it maintain for bugs accounting where each software bug having their report as called bug data. It includes 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. in market, various tools are available for bug triage analysis but they are not fully suitable for large scale and complex data and software repositories. So the software development using data mining techniques to handle software bugs. The data mining techniques helps to handle software repositories such as uncover and hidden information related software repositories. It also solves 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 trigger 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. Bug repository to which both software developers and users can report bugs or defects or issues in the software, suggest possible enhancements, and comment on existing bug reports. The advantage of an open bug repository is that it may allow more bugs to be identified and solved. The first goal of bug management is to manage software bug. Bug fixing process is very important and time consuming process in manual software maintenance. In software development process we faced like large scale databases for storing the output of the software development. In software development process we faced like large scales databases for storing the output of the software development. Bug triage is most efficient process for bug fixing, to allocate a new bug and handle it.

II. LITERATURE SURVEY

Table 1: Literature Survey

Sr no.	Paper	Proposed	Advantages	Disadvantages
1	Towards effective bug triage using software data reduction techniques. [9]	In this paper, we propose the problem of data reduction for bug triage to reduce the scales of data sets and to improve the quality of bug reports. We use technique	<u>Word dimension</u> We use feature selection to remove noisy duplicate words in a data set <u>Bug dimension</u> In stance selection can remove uninformative bug	1. We propose a combination approach to addressing the problem of data reduction. 2. This can be viewed as an application of instance selection and feature selection in bug repositories

		s of instance selection and feature selection to reduce noise and redundancy in bug data sets. However, not all the noise and redundancy are removed.	reports, meanwhile, we can observe that the accuracy may be decreased by removing bug reports. Word dimension : By removing uninformative words, feature selection improves the accuracy of bug triage. This can recover the accuracy loss by instance selection.	.We present the problem of data reduction for bug triage. This problem aims to augment the data set of bug triage in two aspects, namely a) to reduce the scale of bug dimension and word dimension.		identifying noisy instances in defect data			
					3	Improving bug triage with tossing graphs [5].	Author proposed bug tossing graph model can be easily incorporated into existing bug triaging systems.	Proposed the model increased the prediction accuracy by up to 23 percentage points compared to traditional bug triaging approaches. Proposed method is to reduce reassignment in bug triage.	Current model is based on regular Markov chains and thus only use the current state for prediction
					5	Towards more accurate retrieval of duplicate bug reports	Proposed approach is twofold, first BM25F is an effective textual duplicates measure that is designed for short unstructured queries and secondly a new retrieval function REP fully	Improved the accuracy of duplicate bug retrieval.	Need to speed up the retrieval process.
2	Dealing with noise in defect prediction [12]	Author introduced a method to measure noise conflict in software defect prediction and also proposed a new method called CLNI for	Performance and accuracy is improved by using proposed approach.	The limitation of their method is that mislabeled instances are often not outliers.					

		utilizing text and other information available in reports such as product.		
6	SV-kNNC: An algorithm for improving the efficiency of k-nearest neighbor [13].	Paper proposed SV-kNNC approach for data reduction to enhance performance of kNN.	Advantages of SV-kNNC has the ability to reduce data, the classification time required is less and provides best performance because training data are evaluated twice before classification process.	Not all noisy and redundant data is removed by proposed algorithm.

III. PROPOSED APPROACH

Bug triage is a method where all bug issues identified, prioritized and monitoring of this activity. It helps to ensure that the reported issues whether bugs or improvements or a feature request, are properly managed. So, we proposed system bug data set is input of the system and using data mining mechanism like instance selection and feature selection, concurrently with historical bug data sets. We have added a new one features like feedback session. So, it improvement of software bug quality.

System Architecture

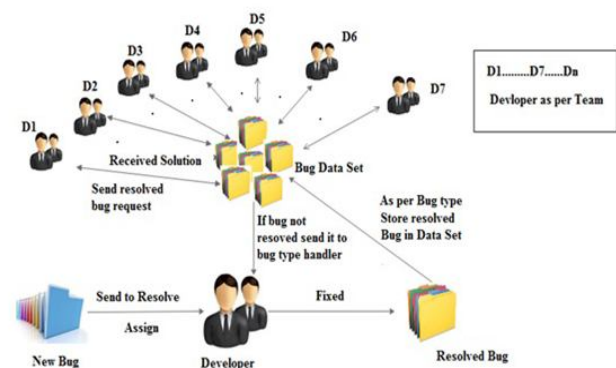


Fig. 1: System Architecture

IV. METHODOLOGY

In this paper, we have survey various methodology for bug triage

1) Instance Selection

Instance selection approaches related with data mining jobs such as classification and clustering .It's a non-trivial procedure of finding usable, new, hypothetically valuable, and at last comprehensible patterns in Original bug Data Reduced

Original bug selecting a subset of records to attain the original determination of a data mining application as if the entire records are used.

The ultimate result of instance selection is mode independent.

$$P (Mw) \cong P (Ms)$$

2) Feature selection

- It selects a minimum set of features such that the probability 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.
- Reduce # of patterns in the patterns, easier to understand.
- Create new attributes that can capture the important information in a data set much more efficiently than the original attributes.
- Use the smallest representation which is enough to solve the task.

3) Heuristic methods

- Step-wise forward selection
- Step-wise backward elimination
- Create new attributes that can capture the important information in a data set much more efficiently than the original attributes
- Three general methodologies:
 - Feature extraction
 - domain-specific
 - Mapping data to new space (see: data reduction)

V. CONCLUSION

Many software industries invest their more of cost in dealing with these bugs' data. As Bug triage is efficient step for software maintenance in both labor cost and time cost. The objective is to correctly identify the new bugs for developers and resolves using various data mining techniques like PSO, RTL, Text Classification algorithms etc. so we focused to reduce the large scale of the bug set and to remove the noisy and redundant bug reports for bug triage. And it improves the less scale of bug data and quality of software.

REFERENCES

- [1] Pankaj Gakare, Yogita Dhole, Sara Anjum3, "Bug Triage with Bug Data Reduction", Volume: 02 Issue: 04 | July-2015, e-ISSN: 2395 -0056 Volume: 02 Issue:04,2015, IRJET p-ISSN: 2395-0072.
- [2] Jayashri Gholap, N. P. Karlekar, "Survey on Bug Triage with Software Data Reduction Techniques", IJSR, ISSN (Online): 2319-7064.
- [3] Kypros Constantinides, Onur Mutlu, and Todd Austin, Online Design Bug Detection: RTL Analysis, Flexible Mechanisms, and Evaluation, in proceedings of the 41st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-41), November 2008.
- [4] G. Jeong, S. Kim, and T. Zimmermann, "Improving 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.
- [5] Q. Shao, Y. Chen, S. Tao, X. Yan, and N. Anerousis, "Efficient ticket routing by resolution sequence mining" ,in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2008, pp. 605– 613
- [6] 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. Eng., 2011, pp. 253–262
- [7] S.Amritha, Ajinnifer Sagaya Rain, "bug triage using dimensionality reuction techniques and PSO algorithm", IJECS, VSI6 jun-2016, p-no.16800-16084.
- [8] Jifeng Xuan, He Jiang, Member, IEEE, Yan Hu, Zhilei Ren, Weiqin Zou, Zhongxuan Luo, and Xindong Wu, Fellow, IEEE, "Towards Effective Bug Triage with Software Data Reduction Techniques", IEEE transactions on knowledge and data engineering, vol. 27, no. 1, January 2015.
- [9] J. Anvik, L. Hiew, and G. C. Murphy, "Who should fix this bug?" in Proc. 28th Int. Conf. Softw. Eng., May 2006, pp. 361–370.
- [10] C. C. Aggarwal and P. Zhao, "Towards graphical models for text processing," Knowl. Inform.Syst., vol.36, no 1, pp. 1–21, 2013.
- [11] S. Kim, H. Zhang, R. Wu, and L. Gong, "Dealing with noise in defect prediction," in Proc. 32nd ACM/IEEE Int. Conf. Softw. Eng., May 2010, pp. 481–490.
- [12] A. Srisawat, T. Phienthrakul, and B. Kijisirikul, "SV-kNNC: An algorithm for improving the efficiency of k-nearest neighbor" in Proc. 9th Pacific Rim Int. Conf. Artif. Intell., Aug. 2006, pp. 975–979.