
Banupriya. A. S¹, Pushpalatha M. N²
Department of Information Science and Engineering
M S Ramaiah Institute of Technology, Bengaluru, India

Abstract:
The development of software project includes many phases. Each phase has its own importance. Software maintenance is essential phase of software development. To improve software reliability, software defect prediction is applied to the process of software maintenance to identify potential bugs. If we categories, the bugs based on severity and priority will help in assigning the bugs to developer. In this paper, we have presented review on the different types of machine learning algorithm, Information retrieval-based approach, emotional based approaches for bug report analysis and proposed the method for predicting the severity and priority of bug report using deep learning and machine learning techniques.

Keywords: Priority, Severity, Software maintenance, Bug triage.

I. INTRODUCTION

Software bugs is a big issue in the development of the software. This problem last for long time and causes the error in the development. These usually created by the programmer’s mistake. There are various reasons for this issue. A bug report contains all the information that relevant. A Bug report is sent to the person in charge of it. The reports are usually submitted to the bug tracking system. Due to this bug repository size keeps on increasing in an extreme rate. Because of the increased size of the software repository it causes the biases in the bug triage process. The bug triage is the process where the tracker issues will be screened and also prioritized. The bug reports will represent the bugs that the users encountered while using the released software systems. Developers will be interested in the collection such bugs in order to improve the software systems. The bug priority is something that should be resolved based their urgency level. The classification of the bug based on priority will tell us which should be resolved at what time.

The classification of the bug priority is as follows:

- Immediate – The bug that must be solved immediately.
- High – The high priority level of the bug must be resolved early in the development of the software so that it will not affect the product after release.
- Medium – The medium priority level of the bug should be resolved after solving the bugs which has high importance.
- Low – This bug can be resolved in a future also or it can be not resoled also.

Severity defines how serious the bug is and it also tells that how badly it will be affected to the functionality. Severity will be predefined field for the Bug. Default severity values are Normal, Blocking, Small, Enhancement, Critical. Blocker severity is actually meaning that the systems or the functionalities will not be currently unavailable because of this bug. The severity level Enhancement will be the lowest level of severity among others used for minor bugs.

II. LITERATURE REVIEW

The literature review gives view of the previous papers on the related topic. Survey made on the topic and find out the discovery happened on the area of the related topic. This section focuses on the literature survey carried out in the field of prediction of priority and severity involved in the bug reports. Few selected research papers related to the proposed technique development and validation has been discussed in detailed.

Guisheng Fan et al.,[1] have proposed a model which learns the automatic semantic information. The model they have proposed is Defect prevention via Recurrent neural network. Here they have make use of the seven java open source projects for the experiment. They have made use of the area under curve and the F1 measures for evaluation. From this they found that, Area under the curve improved by 7% and the F1 measures by the 14%. Author stated that further they want to implement with other projects and with the other programming languages.

Ashima Kukkar et al.,[2] have proposed deep learning model called Bug Severity classification. Here the author make use of the Convolution neural network and they also make use of the Random forest with Boosting. Author found that by applying these models the F1 measures gradually increased. The existing system was giving the F1 measure of 84.24%, after implementing with these techniques its giving about 96.43%. Further author wants to implement with other deep learning techniques.

Ha Manh Tran et al.,[3] have propose a machine learning algorithm for the classification of priority and severity for the datasets. Author made use of all the machine learning algorithm...
for experiment. Further they carried out with the random forest. Author found that random forest gave the 0.75 score and they say that its sufficient for the determination of the features. Further mentioned that they want to improve the accuracy with random forest. They also mentioned that decision tree takes less processing time when it is compared with the decision tree.

Waheed yousuf ramay et al.,[4] proposed a model called deep neural network, an automatic approach for the prediction severity in the bug report. Author found that, after implementing with the convolution neural network, the accuracy increased by 7.90%. They have mentioned about different machine learning and deep learning techniques so far applied by different authors in their papers.

Senthil Mani et al.,[5] have proposed a model called deep bidirectional recurrent neural network with attention, which represent the bug report and learns with semantic feature and syntactic from bigger word in an unsupervised manner. By implementing this model, they found that model could remember the important text parts in the bug report.

Meera Sharma et al.,[6] have proposed a model they evaluated with the different machine learning algorithm like naïve bayes, Neural network, Support vector machine, K nearest neighbor. Author found that implementing with all these found that accuracy in predicting the priority was more than 70% except for the Naïve bayes.

Qasim umer et al.,[7] have proposed model which predicts the priority of the bug report using the emotional based model which automatically assign to predict the level of bug priority. Author have mentioned the steps how the prediction is done. They made use of the machine learning algorithm and found that f1 measures increased by 6%.

Jifeng Xuan et al.,[8] have proposed a model that developer prioritization in bug repositories by extending a socio-technical approach. Author analyze three problems of the developer prioritization, namely the characteristics in products, the evolution, and the tolerance of noises. Based on the analysis, they investigate the ways to leverage the developer prioritization to improve three typical tasks in bug repositories. The results are studied on over 900000 bug reports in Eclipse and Mozilla.

Mamdouh Alenezi et al.,[9] have proposed a model that predicts the bug priority. Author implemented using different machine learning algorithm like Random forest, Decision tree and Naïve bayes. Author make use of the firebox and the eclipse for the experiment. They found that in terms of accuracy decision tree and random forest was better than the naïve bayes.

Menzies et al.,[10] have developed an automated method called SEVERIS(SEVERity Issue assessment) for hindering the test engineer in allocating severity levels for defect reports. Their proposed approaches used text mining and machine learning approaches applied on the conventional existing defect reports. Authors has also presented case study of the proposed method by using datasets from NASA’s Project and Issue Tracking System (PITS).

Ahmed Lamkanfi et al.,[11] have proposed invigilation how the severity of the bug report can be predicted using the text description. They have made use of the text mining on the three open source i.e, GNOME, Mozilla, Eclipse. They found that performance has been increased with the precision and recall.

K. K. Chaturvedi et al.,[12] have proposed the machine learning techniques like k-nearest neighbor, Naïve bayes, Naïve bayes multinomial, support vector machine, J48 and Ripper in order to predict the severity of bug report. They found that best f-measure was found for the severity level 2,3,4.

Fareen Sayed et al.,[13] have proposed the NB classifiers where we can predict that which bug can be allocated to appropriate developer. Here they found that issue of data reduction has been removed by making use of the techniques like instance selection and keyword selection.

Shreya Arudkar et al.,[14] have proposed the model which predict that which of the developer should be assigned with the which type of the bug by using the textual description of the bug. They have made use of the machine learning technique i.e, Random forest on the two large projects like mozilla and eclipse and they found increase the accuracy of the tracking system.

John Anvik et al.,[15] have proposed a model of machine learning techniques which automatically suggest the developer to which it has to be assigned. They made this experiment on the two large projects i.e, Eclipse and Firefox.

He Jiang et al.,[16] have proposed a technique which decrease the scale of the bug dataset. Here they have combine the instance selection and future selection to improve the quality of data. They applied these techniques on Large open source projects i.e, Mozilla and eclipse. In determining to apply the instance and future selection order, they made use of the historical data of the particular datasets.

Salma Shaik et al.,[17] have proposed a model which is the combination of the instance selection and future selection which will increase the quality of the data and decreases the level of the bug datasets. Here they have also found that there will be access restriction for the developers so that other developer cannot access it.

Philipp Schugerl et al.,[18] have proposed the techniques like natural language processing and Information retrieval. They also made use of the naïve bayes machine learning algorithm for their project. They found that classifying was easier for the good quality of the description of the bug.

Dr. Arvinder Kaur et al.,[19] have proposed a tool, from the bug repository data will be extracted using REST API. This implemented using the c#. Here they made use of the Jira and Bugzilla. They have implemented the automated collection of bug report from the bug repository which will be helpful in the prediction of priority and for classification and many more.

M Chalapathi Rao et al.,[20] have proposed method to look into the bug report and evaluate and classify the severity. Here they make use of the Software Bug Complexity Cluster(SBCC) using Self Organizing Maps (SOM) approach. They found that 85% more severity bug has been found using SOM than k-means.

Naresh Kumar Nagwani et al.,[21] have proposed a new algorithm in their project. This algorithm helps in identifying the
developer for the bug report. It identifies the developer who is expert for that particular reported bug. Here they look into the textual information of the bug, then they will identify the repeated word and find the expert developer.

Ashok Bhansali et al.,[22] have proposed a model where the complexity of the bug is predicted. They applied clustering technique on the software repositories so that the complexity of the cluster is created. They applied on this open source bug repository of mysql.

Hadi Hosseini et al.,[23] have proposed mechanism to assign the bug to the developer. This mechanism is called as auction based multiagent. They used data mining technique in their project for the prediction. They found that 16% increment in fixing up the bug compared to earlier data.

Swati Jain et al.,[24] have proposed model which is used to remove the unwanted information so that the quality of the data will be increased. They mentioned that if theses non informative information are removed will give the better results. They made use of the Naïve Beyer method for the classification. They also mentioned that they made use of the feature selection and instance selection.

Anjali Goyal et al.,[25] have proposed a model called DevNet, its heterogenous network of the developer collaboration framework. It will represent the developer collaboration and also analyse in the bug repositories. They mentioned that from the experimental the accuracy increased from 5 to 15% in the bug triage. They carried out the experiment on the large open source projects like eclipse and Mozilla.

Hao Hu et al.,[26] have proposed a model a called BugFixer. These methods actually tell the which developer is suitable for that particular bug to resolve from the previous history. Developer-component-bug (DBC) is constructed by the BugFixer which captures the previous resolved bug history of developer. They carried out the experiment on the three large projects like Eclipse, Mozilla, NetBeans and also on other two small projects. They found that their model gave better results for the large three projects compared to smaller one.

Deqing Wang et al.,[27] have proposed a model Rebug detector which make use of the code feature and also the information the bug to detect the related bugs. They made use of the open source project Apache Lucene-Java where they applied the Rebug detector. They found how their tool helped in finding the real bug in the project.

Liu Yu et al.,[28] have proposed a algorithm which is used to give the software fault case from the bug report of the software. They mentioned that they carried out this project on large three open source projects like Firefox, Apache, Eclipse. They found that from this reduction of time and as well as the cost because no need to build the fault case from the beginning.

Chen Liu et al.,[29] have proposed an approach R2 Fixer, this used to generate the patches for bug fixing. The semantic path recognition, bugs which are previously fixed, and the machine learning techniques are the combination of the R2Fixer. They have mentioned that they carried out this project on the apache, Linux kernel, Mozilla.

Davor Cubranic et al.,[30] have proposed a model which will automatically assign the bug to the developer. They made use of the naïve bayse classification algorithm for the text classification on the open source project Eclipse. They mentioned that results were 30% in the assignment of bug to the developer.

III. CONCLUSION

Bug triage will assign the bugs to the developer who will have the knowledge in solving the bug or have experience in solving the related bug. This helps in reducing time consumption and also the cost. In this paper, we have mentioned different types of machine learning algorithm, Information retrieval-based approach, emotional based approaches. By this survey it is observed different methods can be used for bug triage.

IV. REFERENCES


[24]. Swati Jain, Swapna Rose Wilson, “AUTOMATED BUG ASSORTMENT SYSTEM IN DATASETS”.


[30]. Davor Cubranic, Gail C. Murphy, “Automatic bug triage using text categorization”.