© 2016 IJSRSET | Volume 2 | Issue 3 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

An Efficient Approach to Predict Software Defect - Review

Vanita Patni*, Divyesh Joshi

Department of Computer Engineering, PIET, Baroda, Gujarat, India

ABSTRACT

Faults in software systems are a major problem. A software fault is a defect that causes software failure in an executable product. Quality of a product is correlated with the number of defects as well as it is limited by time and by money. The possibility of early estimating the potential faultiness of software could help on planning, controlling and executing software development activities. So, defect prediction is very important in the field of software quality and software reliability. With the increase of the web software complexity, defect detection and prevention have become crucial processes in the software industry. It is applied to web based systems using graph-based clustering algorithms. An appropriate implementation of the graph-based clustering in defect prediction may facilitate to estimate defects in a web page source code.

Keywords: Software defect prediction, graph-based clustering

I. INTRODUCTION

With the growing up of the World Wide Web and its applications, web based applications provide a wide range support to various platforms. This diversity yields a number of product development processes as a result of web market competition. On the other hand, software quality assuring is gradually becoming harder with the shortened development processes in this competition. As the software project's scale expands software test process time increases in parallel. Traditional test approaches are becoming insufficient across increasing complexity of web based applications. The main objective of the web application testing is to detect and fix misbehavior, properties of system according to specific input values. Clustering is a non-hierarchical procedure in which items are moved among sets of clusters until the desired set is reached. Each clustering technique makes some assumptions about the underlying data set. If the assumptions hold, good clustering can be expected. But, it is hard to satisfy all the assumptions. Therefore, it is beneficial to apply different clustering methods on the same data set, or the same method with varying input parameters or both. Graph-based clustering is flexible structure. The usage of graph based algorithm could ease to specify defect distribution. Its computation is fast and scalable.

II. LITERATURE SURVEY

2.1. A novel defect prediction method for web pages using k-means++, Muhammed Maruf Ozturk, Unal Cavusoglu, Ahmet Zengin, ELSEVIER 2015 [1].

In this paper, they use a new defect clustering method using k-means++ for web page source codes. According to the experimental results, almost half of the defects are detected in the middle of web pages. k-means++ is significantly better than the other four clustering algorithms in three criteria on four data set. They also tested their method on four classifiers and the results have shown that after the clustering, Linear Discriminate Analysis is, in general, better than the other three classifiers.

TABLE I

COMPARISON OF CLUSTERING ALGORITHM [1]

Algorithm	Evaluation
K Means	Less iteration and time, good result
K-Means++	More iteration and time, best result

Fuzzy-k	More iteration and time, unstable result
SPSS	Less iteration and time, good result

2.2. Extracting software static defect models using data mining, Ahmed H. Yousef, Elsevier , 2015[2].

In this paper, a data mining approach is used to show the attributes that predict the defective state of software modules. Software solution architecture is proposed to convert the extracted knowledge into data mining models that can be integrated with the current software project metrics and bugs data in order to enhance the prediction. The results show better prediction capabilities when all the algorithms are combined using weighted votes.

When only one individual algorithm is used, Naive Bayes algorithm has the best results, then the Neural Network and the Decision Trees algorithms.

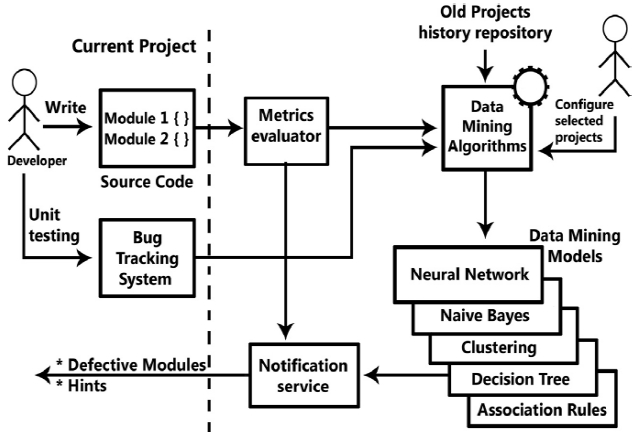

Figure 1. Architecture for defective module identification using data mining and Software metrics [2]

2.3. Graph Based Analysis and prediction for software Evolution, Pamela Bhattacharya, Marios Iliofotou, Iulian Neamtiu, Michalis Faloutsos, IEEE 2012[3]

This paper provided a graph construction method and a set of metrics that capture the structure and evolution of software products and processes. Source code-based graph metrics can reveal differences and similarities in structure and evolution across programs, as well as point out significant events in software evolution that other metrics might miss. Also shown that graph metrics can be used to predict bug severity, maintenance effort and defect-prone releases.

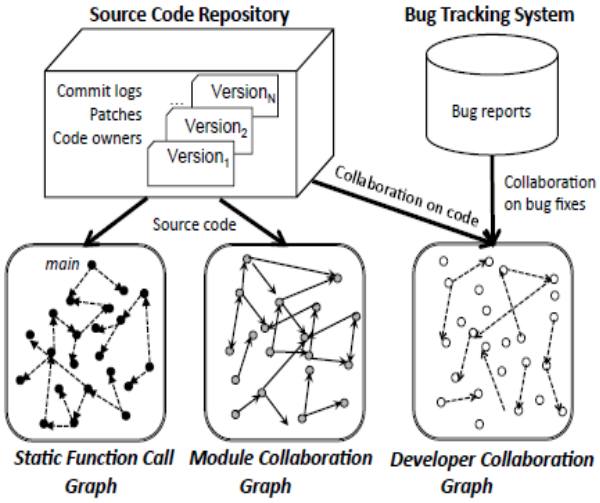

Figure 2. System Overview[3]

2.4. Open Issue in Software Defect Prediction, Ishani Aroraa, Vivek Tetarwala, Anju Sahaa, Elsevier, 2014[4]

In this paper, an aggregate of six problems was discussed: finding the set of attributes to be correlated with fault, the absence of standard measures for performance assessment, problems with cross project defect prediction, economics of software defect prediction and class imbalance problem and the absence of any general framework for the software defect prediction.

2.5. A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks, Journal of Software Engineering, April 2015[5]

In this paper, proposed some techniques for improving the accuracy of machine learning classifier for software defect prediction by ensemble some machine learning methods, by using boosting algorithm, by adding feature selection and by using parameter optimization for some classifiers. The results of this research also identified three frameworks that are highly cited and therefore influential in the software defect prediction field.

III. CONCLUSION

Defects can assess in directing the software quality assurance measures as well as improve software management process if developers find and fix them early in the software life cycle. Effective Defect prediction is based on good data mining model. In this surveyed different data mining algorithms used for defect prediction. To select a better data mining algorithm, domain expert must consider the various factors like problem domain, type of data sets, nature of project, uncertainty in data set etc.

IV.REFERENCES

[1] Muhammed Maruf Ozturk, Unal Cavusoglu, Ahmet Zengin, A novel defect prediction method for web pages using k-means++, ELSEVIER 2015

[2] Ahmed H. Yousef, Extracting software static defect models using data minig, Elsevier , 2015.

[3] Pamela Bhattacharya, Marios Iliofotou, Iulian Neamtiu, Michalis Faloutsos,Graph Based Analysis and predictionforsoftwareEvolution,IEEE2012

[4] Ishani Aroraa, Vivek Tetarwala, Anju Sahaa , Open issues in software defect prediction,Elsevier, 2015

[5] A Systematic Literacture Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks,JournalOfSoftwareEngineering,April 2015.

[6] Golnoush Abaei · Ali Selamat, A survey on software fault detection based on different prediction approaches, springer, 2014

[7] Luis A. Leiva, Enrique Vidal, Warped K-Means: An Algorithm to cluster sequentially-distributed data, Elsevier, 2013

[8] Bahman Bahmani, Benjamin Moseley, Andrea Vattani, Ravi kumar, Sergei Vassilvitskii, Scalable K-Means++, VLDB Endowment, 2012

[9] Ali Hadian, Saeed Shahrivari,High performance parallel k-means clustering for disk-resident datasets on multi-core CPUs, Springer, 2014

[10] Hesham Abandah, Izzat Alsmadi, Call Graph Based Metrics To Evaluate Software Design Quality, International journal of software Engineering and Its Applications, January 2013.