IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 50-54

Analysis of various fault prediction models for software and evaluation of its performance, validation

Miss. S. V. Athawale[1], Dr. V. M. Thakare [2]
Student in Department of Computer Science and Information Technology, SGBAU, Amravati, Head of Department of Computer Science and Information Technology, SGBAU, Amravati.

Abstract: *Software systems are being utilized to solve or model increasingly sophisticated and complex problems in a variety of application domains. Software fault prediction has been one of the active parts of software engineering, but to date, there are few test cases prioritization technique using fault prediction. There are many models are developed for fault prediction in software development lifecycle. In this paper the analysis of various fault prediction models is proposed. It analyses various software faults prediction models and does the evaluation of its performance, validation.*
Keywords: *Software fault prediction, feature selection, classification, software metric selection*

I. Introduction

Software fault prediction (SFP) is useful for helping the software engineer to locate potential faulty modules in software testing more easily, so that it can save a lot of time and budgets to improve the software quality. There are various software fault prediction models are developed using various methods. Each software fault prediction models have their own application with different behavior to perform their task. This paper, discusses five different software fault prediction approach such as Software Fault Prediction Based On One-Class SVM [1],An Investigation of Essential Topics on Software Fault-Proneness Prediction [2], Clustering Approach Combining Fault Prediction For Test Case Prioritization [3],Software Defect Prediction based on Geometric Mean for Subspace Learning [4] and Artificial neural network-based metric selection for software fault-prone prediction model [5]. These software fault prediction models analysis gives performance, validation of various SFP models based on various parameters.But these methods also have some problem so to overcome such problems "Analysis of various fault prediction models for software and evaluation of its performance, validation" is proposed here that is based on parameters.

II. Background

Many studies on software fault prediction models have been done to develop the software fault prediction approach in recent past years. Such approaches are:

Clustering Approach Combining Fault Prediction For Test Case Prioritization is proposed to implement a new prioritization technique incorporating clustering algorithm and code fault prediction [1]. SFP Based On One-Class SVM that aiming at solving the problem that the faulty samples are too rare to train a classifier [2]. Software fault Prediction based on Geometric Mean for Subspace Learning, isa method based on dimensionality reduction technique and CRF Model is introduced to improve software defect prediction in imbalance distribution by choosing the best combination of features for data set and incorporating complex features without any process in training stage. [3].An Investigation of Essential Topics on Software Fault-Proneness Prediction that represents an investigation of essential topics in this area, including techniques for evaluating the effectiveness of fault-proneness prediction models[4].Artificial neural network-based metric selection for software fault-prone prediction model propose a reduction dimensionality phase, which can be generally implemented in any software fault-prone prediction model [5]. This paper introduces five software fault prediction i.e. An Empirical Study On Clustering Approach Combining Fault Prediction For Test Case Prioritization, Software Defect Prediction based on Geometric Mean for Subspace Learning, Software Fault Prediction Based On One-Class SVM, An Investigation of Essential Topics on Software Fault-Proneness Prediction, Artificial neural network-based metric selection for software fault-prone prediction model.These are organized as follows. **Section I** Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on software fault prediction models. **Section VI** proposed method and outcome result possible. Finally **section VII** Conclude this review paper.

III. Previous Work Done

In research literature, many software fault prediction models have been studied to provide various fault prediction schemes and improve the performance in terms of accuracy, effectiveness, precision.

Lei Xiao et al. (2017) [1] have worked on Clustering Approach Combining Fault Prediction For Test Case Prioritizationthat uses clustering algorithm which have the best cluster number to test case prioritization which improve the efficiency of regression testing.

Yan Gaoet al. (2017) [2] has proposedSoftware Defect Prediction based on Geometric Mean for Subspace Learning usesGMCRF (Geometric Mean Conditional Random Field) method which achieves much better final results than the other approach.

Lin Chenet al. (2016) [3] has worked on One-Class SVMfor Software Fault Prediction thataiming at solving the problem that the faulty samples are too rare to train a classifier. One-class is a relatively special training method in machine learning.

Shou-Yu Lee et al. (2016) [4] presents An Investigation of Essential Topics on Software Fault-Proneness Prediction in which they offer a fast and convenient index not only for professionals to efficiently identify their issues of concern.

C. Jinet al. (2012) [5] have proposed Artificial neural network-based metric selection for software fault-prone prediction that presents applications of artificial neural network (ANN) and support vector machine in software fault-prone prediction using metrics.

IV. Existing Methodologies

Many software fault prediction approach have been implemented over the last several decades. There are different methodologies that are implemented for different software fault prediction models i.e. Clustering Approach Combining Fault Prediction For Test Case Prioritization, Geometric Mean for Subspace Learning, One-Class SVM, An Investigation of Essential Topics on Software Fault-Proneness Prediction, Artificial neural network-based metric selection for software fault-prone prediction model.

An Empirical Study On Clustering Approach Combining Fault Prediction For Test Case Prioritization: Clustering Approach Combining Fault Prediction for Test Case Prioritizationimplement a new prioritization technique incorporating clustering algorithm and code fault prediction. Author investigate adopting clustering algorithm which have the best cluster number to test case prioritization whether could improve the efficiency of regression testing.In this both test cases prioritization in each cluster and cluster prioritization considered, the cluster prioritization depends on the result code fault prediction based SVM algorithm, and test cases in each cluster prioritization depend on the distance between test case and center point [1].

Software Defect Prediction based on Geometric Mean for Subspace Learning: In this Software Defect Prediction based on Geometric Mean for Subspace Learning model,GMCRF (Geometric Mean Conditional Random Field) method based on dimensionality reduction technique and CRF Model is introduced to improve software defect prediction in imbalance distribution by choosing the best combination of features for data set and incorporating complex features without any process in training stage. GMCRF can easily be trained using the simple direct optimization technique of stochastic gradient descent. This GMCRF method achieves much better final results than the other approach[2].

Software Fault Prediction Based On One-Class SVM: Software Fault Prediction Based On One-Class SVM, aims at solving the problem that the faulty samples are too rare to train a classifier, an one-class SFP model is proposed by using only non-faulty samples based on one-class SVM. The empirical validation is conducted on 6 extremely imbalanced datasets collected from real-world software containing only small amounts of faulty instances. The test results suggest that the proposed model can achieve a reasonable fault prediction performance when using only a small proportion of training sample and performs much better than conventional and class imbalanced learning based SFP models in terms of G-mean measure [3].

An Investigation of Essential Topics on Software Fault-Proneness Prediction: An Investigation of Essential Topics on Software Fault-Proneness Prediction that represents an investigation of essential topics in this area, including techniques for evaluating the effectiveness of fault-proneness prediction models, issues of concerns when building the prediction models, as well as findings shared by the academic community. In order to build a fault-proneness prediction model, a multitude of factors need to be considered during model construction. In order to reduce the risk of software faults manifesting during operation, techniques are needed to identify code which has the potential to cause problems early on so that more effort can be spent on testing to prevent such problems from occurring. This is the reason that stimulates the proposal of various fault-proneness prediction models. Author presented an objective to offer a fast and convenient index not only for professionals to efficiently identify their issues of concern in software fault-proneness prediction but also for researchers [4].

Artificial neural network-based metric selection for software fault-prone prediction model: Artificial neural network-based metric selection for software fault-prone prediction model, authors proposed a reduction

dimensionality phase, which can be generally implemented in any software fault-prone prediction model. In this, the authors present applications of artificial neural network (ANN) and support vector machine in software fault-prone prediction using metrics. A new evaluation function for computing the contribution of each metric is also proposed in order to adapt to the characteristics of software data. The vital characteristic of this approach is the automatic determination of ANN architecture during metrics selection. It is also very simple because its implementation requires neither extra cost nor expert's knowledge. The proposed model has good performance, and can provide software project managers with trustworthy indicators of fault prone components [5].

V. Analysis And Discussion

An Empirical Study on Clustering Approach Combining Fault Prediction for Test Case Prioritization discuss a test case technique adopting clustering algorithm and fault prediction which acquire the best number of clusters by DB index[1].Software Defect Prediction based on Geometric Mean for Subspace Learningshows that GMCRF can easily be trained using the simple direct optimization technique of stochastic gradient descent. The proposed method involves dimensionality reduction and prediction technique [2]. Software Fault Prediction Based on One-Class SVM shows anone-class SFP model which is proposed by using only non-faulty samples based on one-class SVM. The proposed model can facilitate the SFP with few faulty modules in early life of software testing [3]. An Investigation of Essential Topics on Software Fault-Proneness Predictionshows investigation of essential topics, including techniques for evaluating the effectiveness of fault-proneness prediction models, issues of concerns when building the prediction models [4]. Artificial neural network-based metric selection for software fault-prone prediction model,present applications of artificial neural network (ANN) and support vector machine in software fault-prone prediction using metrics. It is very simple model because its implementation requires neither extra cost nor expert's knowledge [5].

Table-1: Comparisons between different software prediction approaches

Prediction models and approach	Advantages	Disadvantages
Clustering Approach Combining Fault Prediction for Test Case Prioritization	This approach improves the rate of fault detection of test cases.	This approach has high cost and the time required for clustering algorithm is more than other methods.
Geometric Mean for Subspace Learning	This method can easily be trained using the simple direct optimization technique of stochastic gradient descent. This GMCRF method achieves much better final results than the other approach.	In this, not all dimensionality reduction methods are suitable to select feature for SDP task.
One-Class SVM	This method achieve a reasonable fault prediction performance when using only a small proportion of training sample and performs much better than conventional and class imbalanced learning based SFP models in terms of G-mean measure.	This method need to investigate more imbalanced learning methods to verify the proposed model.
Investigation of Essential Topics on Software Fault-Proneness	This method provides convenient index to efficiently identify their issues of concern in software fault-proneness prediction.	Drawback is that techniques are constantly improving, software, too, is becoming increasingly more complex.
Artificial neural network-based metric selection	This model has good performance, and can provide software project managers with trustworthy indicators of fault prone components.	This model need to conduct additional experiments to tackle common problems that affect the training/test results.

VI. Proposed Methodology

Software fault prediction in software projects improves the software productivity. So quality assurance of software, accuracy of result, achievable performance and effectiveness are very important in this nature. There are still some problems which trouble in this field. New software fault prediction method called "Analysis of software fault prediction models and evaluation of its performance, validation" for more effective and more accurate fault prediction model is propose here to analyses the various models. As this analysis is based on the various parameters, which would evaluate the performance, validation of fault prediction.

One of the most important goals of such fault prediction techniques is to accurately predict the modules where faults are likely to hide as early as possible in the development lifecycle. Software fault prediction (SFP) is useful for helping the software engineer to locate potential faulty modules in software testing more easily, so

that it can save a lot of time and budgets to improve the software quality.In this proposed paper, first analysis of various software fault prediction models is performed, and then evaluates the performance, validation of each model according to various parameters like accuracy, precision, f-measures, recall, etc. Analysis of various software fault prediction models is proposed to improve and evaluate the performance, validation of existing results and it is necessary to validate the developed fault prediction model on different data set from which they are trained. Analysis is the process of analyzing and investigating the performance of models, issues of concerns when building the prediction models. Validation is a well-known technique in software testing field which is used to evaluate and compare predictive models. An objective is to offer a fast and convenient index to efficiently identify their issues of concern in software fault-proneness prediction. WEKA tool is used here to perform the software fault prediction which is simple and easy to use.
Diagrammatic representation of proposed method is shown as follows:

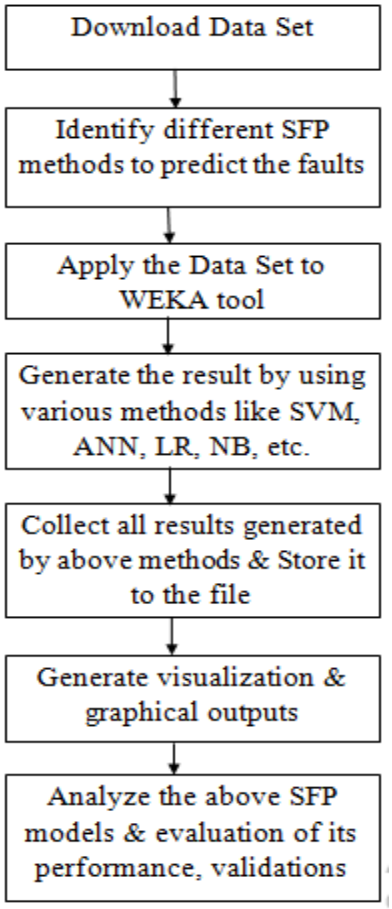

Figure-1: *Workflow of proposed method for software fault prediction*

VII. Outcome And Possible Result

In this way the proposed method is performed for the analysis of various fault prediction models. With the help of analyzing various SFP models, a fast and convenient index to efficiently identify their issues of concern in software fault-proneness prediction is obtained. The main purpose of analyzing software fault prediction method is to evaluate the performance and validation of such methods with the improved parameter of fault prediction model for software. The proposed method gives effectiveness and accuracy of software fault prediction model and improves the performance of fault prediction.

VIII. Conclusion

This paper focused on the study of various software fault prediction approaches i.e. An Empirical Study On Clustering Approach Combining Fault Prediction For Test Case Prioritization, Software Defect Prediction based on Geometric Mean for Subspace Learning, Software Fault Prediction Based On One-Class SVM, An Investigation of Essential Topics on Software Fault-Proneness Prediction, Artificial neural network-based metric selection for software fault-prone prediction model. But there are some problems in accuracy and

performance of software data so to improve this "Analysis of various fault prediction models for software and evaluation of its performance, validation" is proposed here. Analysis of software fault prediction models is proposed to improve and evaluate the performance of existing results that gives fast and convenient index to efficiently identify their issues of concern in software fault-proneness prediction.

IX. Future Scope

From observations of the proposed method the future work will include exact accuracy and high performance of software fault prediction with the help of more performance parameters.

References

[1]. Lei Xiao, Huaikou Miao, Weiwei Zhuang, Shaojun Chen, "An Empirical Study On Clustering Approach Combining Fault Prediction For Test Case Prioritization", IEEE Computer Society,5090-5507, *Pg No.* 815 - 820, *MAY 2017.*

[2]. Yan Gao, Chunhui Yang, Lixin Liang, "Software Defect Prediction based on Geometric Mean for Subspace Learning",*IEEE,* 4673-8979, *Pg No.* 225-229, AUGUST 2017.

[3]. Lin Chen, Bin Fang, Zhaowei Shang, "Software Fault Prediction Based On One-Class SVM", International Conference on Machine Learning and Cybernetics, 5090-0390, *Pg No*.1003-1008, *JULY, 2016.*

[4]. Shou-Yu Lee, Dong Li, Yihao Li, "An Investigation of Essential Topics on Software Fault-Proneness Prediction", International Symposium on System and Software Reliability (ISSSR), 5090-5563, *Pg No. 113-119, AUGUST, 2016.*

[5]. C. Jin, S.-W. Jin, J.-M. Ye, "Artificial neural network-based metric selection for software fault-prone prediction model", IET Software,Vol. 6, Iss. 6, *Pg No.* 479–487, *MAY 2012*.