International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

# Classification Rule Mining Based Software Defect Prediction

**Chanderpal Bhardwaj**
M.Tech Scholar Computer Science and Engineering Dept of CSA Ch. Devi Lal University ,Sirsa(Haryana)
Chanderpal2012@gmail.com
**Dr. Kapil kaswan**
Assit. Professor Dept of CSA Ch. Devi Lal University, Sirsa(Haryana)
kapilkaswan@gmail.com

## Abstract:

*There has been rapid growth of software development. Due to various causes, the software comes with many defects. In Software development process, testing of software is the main phase which reduces the defects of the software. If a developer or a tester can predict the software defects properly then, it reduces the cost, time and effort. In this paper, we show a comparative analysis of software defect prediction based on classification rule mining. We propose a scheme for this process and we choose different classification algorithms. Showing the comparison of predictions in software defects analysis. This evaluation analyzes the prediction performance of competing learning schemes for given historical data sets (NASA MDP Data Set). The result of this scheme evaluation shows that we have to choose different classifiers rule for different data set.*

**Keywords:** Software defect prediction; classification Algorithm; Confusion matrix.

## Introduction:

There has been a huge growth in the demand for software quality during recent ages. As a consequence, issues are related to testing, becoming increasingly critical. The ability to measure software defect can be extremely important for minimizing cost and improving the overall effectiveness of the testing process. The major amount of faults in a software system is found in a few of its components. Although there is variety in the definition of software quality, it is truly accepted that a project with many defects lacks the quality of the software. Knowing the causes of possible defects as well as identifying general software process areas that may need attention from the initialization of a project could save money, time and working effort. The possibility of early estimating the probable faultiness of software could help on planning, controlling and executing software development activities. A low cost method for defect analysis is learning from past mistakes to prevent future ones. Today, there exist several data sets that could be mined in order to discover useful knowledge regarding defects.

Using this knowledge one should ideally be able to:

a. Identify potential fault-prone software.
b. Estimate the distinct number of faults,
c. Discover the possible causes of faults.

## Related Works

### Regression via classification

In 2006, Bibi, Tsoumakas, Stamelos, Vlahavas, apply a machine learning approach to the problem of estimating the number of defects called Regression via Classification (RvC) [4].The whole process of Regression via Classification (RvC) comprises two important stages:

a) The discretization of the numeric target variable in order to learn a classification model,
b) the reverse process of transforming the class output of the model into a numeric prediction.

### Static Code Attribute

International Journal of Research
International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

Menzies, Greenwald, and Frank (MGF) [5] published a study in this journal in 2007 in which they compared the performance of two machine learning techniques (Rule Induction and Naive Bayes) to predict software components containing defects. To do this, they used the NASA MDP repository, which, at the time of their research, contained 10 separate data sets.

### ANN

In 2007, Iker Gondra [6]used a machine learning methods for defect prediction. He used Artificial neural network as a machine learner.

### Embedded software defect prediction

In 2007, Oral and Bener [7] used Multilayer Perception (MLP), NB, VFI(Voting Feature Intervals) for Embedded software defect prediction. There they used only 7 data sets for evaluation.

### Association rule classification

In 2011 Baojun, Karel [3] used classification based association rule named CBA2 for software defect prediction. In these research they used association rule for classification. and they compare with other classification rules such as C4.5 and Ripper

### Defect-proneness Prediction framework

In 2011, Song, Jia, Ying, and Liu propased a general framework for software defect-pronness prediction. in this research they use M*N cross validation with the dataset(NASA, Softlab Dataset) for learining process. and they used 3 classification algorithms(Naive baysed, OneR, J48). and they copared with MGF [5] framework.
In 2010 a research has benn done by Chen, Sen, Du Ge, [8] on software defect prediction using datamining. In this research they used probabilistic Relational model and Baysean Network.

## Research Objective

Keeping the research indications in view, it has been realized that there exists enough scope to improve the software defect prediction. In this research the objectives are confined to the followings: |

i. To utilize novel data set filtering mechanism for effective noise remove.

ii. To utilize novel classification algorithm for better prediction.

iii. To use better evaluation measurement parameter to get better result.

iv. To decrease the software development cost, time and effort.

## Methodology

- Defect Prediction as a Classification Problem

Software defect prediction can be viewed as a supervised binary classification

problem [2] [3]. Software modules are represented with software metrics, and

are labelled as either defective or non-defective. To learn defect predictors, data

tables of historical examples are formed where one column has a boolean value for

"defects detected" (i.e. dependent variable) and the other columns describe software

characteristics in terms of software metrics (i.e. independent variables).

- Binary classification

1. Bayesian Classification
2. Rule-Based Classification
3. Logistic Regression
4. Decision Tree classification

## CONCLUSION

In order to improve the efficiency and quality of software development, we can make use of the advantage of data mining to analysis and predict large number of defect data collected in the software development.

Clustering based classification can be used.

• Future studies could focus on comparing more classification methods and improving association rule based classification methods

• Furthermore, the pruning of rules for association rule based classification methods can be considered.

**References**

[1] Tao Xie, Suresh Thummalapenta, David Lo, and Chao Liu. Data mining for software engineering. Computer, 42(8):55{62, 2009.

[2] Qinbao Song, Zihan Jia, Martin Shepperd, Shi Ying, and Jin Liu. A general software defect-proneness prediction framework. Software Engineering, IEEE Transactions on, 37(3):356{370, 2011.

[3] Ma Baojun, Karel Dejaeger, Jan Vanthienen, and Bart Baesens. Software defect prediction based on association rule classi_cation. Available at SSRN 1785381, 2011.

[4] S Bibi, G Tsoumakas, I Stamelos, and I Vlahavas. Software defect prediction using regression
via classi_cation. In IEEE International Conference on, pages 330{336, 2006.

[5] Tim Menzies, Jeremy Greenwald, and Art Frank. Data mining static code attributes to learn defect predictors. Software Engineering, IEEE Transactions on, 33(1):2{13, 2007.

[6] Iker Gondra. Applying machine learning to software fault-proneness prediction. Journal of Systems and Software, 81(2):186{195, 2008.

[7] Atac Deniz Oral and Ay_se Ba_sar Bener. Defect prediction for embedded software. In Computer and information sciences, 2007. iscis 2007. 22nd international symposium on, pages 1{6. IEEE, 2007.