On Strategies for Improving Software Defect Prediction

Rahul Krishna, *Dept. of Electrical and Computer Engineering*
North Carolina State University, Email: rkrish11@ncsu.edu

Abstract—Programming inherently introduces defects into programs, as a result software systems can crash or fail to deliver an important functionality. It is very important to test a software throughly before it can be use. But an extensive testing can be prohibitively expensive or may take too much time to conduct This necessitates the use of automated software defect prediction tools. Although numerous machine learning algorithms are available to detect defects in software, but several factors undermine the accuracy of such algorithm. This paper uses Classification and Regression Trees (CART) and Random Forests to examines two approaches to counter the aforementioned problem. The first approach involves the use Synthetic Minority Oversampling Technique (also known as SMOTE). The second approach attempts to use a metaheursitic algorithm such as differential evolution to find the right set of parameters that can change the performance of the predictor.

Index Terms—Defect Prediction, Machine Learning, Differential Evolution, CART, Random Forest.

1 Introduction

2 Results

(a) Ant

Rank	Treatment	Med	IQR
1	RF	41.0	3.0
2	CART	44.0	3.0
3	CART (SMOTE)	70.0	2.0
4	RF (SMOTE)	78.0	1.0

(b) Camel

Rank	Treatment	Med	IQR
1	RF	39.0	1.0
2	CART	43.0	2.0
3	CART (SMOTE)	56.0	2.0
4	RF (SMOTE)	60.0	2.0

(c) Ivy

Rank	Treatment	Med	IQR
1	RF (SMOTE)	0.0	0.0
1	CART (SMOTE)	15.0	15.0
2	RF	50.0	1.0
3	CART	56.0	1.0

(d) Jedit

Rank	Treatment	Med	IQR
1	RF	0.0	0.0
1	CART (SMOTE)	84.0	1.0
1	RF (SMOTE)	88.0	1.0
1	CART	93.0	0.0

(e) POI

Rank	Treatment	Med	IQR
1	CART	36.0	3.0
1	RF	40.0	4.0
2	RF (SMOTE)	53.0	6.0
2	CART (SMOTE)	54.0	4.0

(f) Log4j

Rank	Treatment	Med	IQR
1	RF (SMOTE)	2.0	2.0
2	CART (SMOTE)	14.0	5.0
3	RF	22.0	2.0
4	CART	41.0	2.0

(g) Lucene

Rank	Treatment	Med	IQR
1	CART	47.0	1.0
2	RF	51.0	1.0
2	CART (SMOTE)	50.0	4.0
3	RF (SMOTE)	56.0	3.0

(h) PBeans

Rank	Treatment	Med	IQR
1	RF	51.0	0.0
1	CART	53.0	0.0
1	CART (SMOTE)	56.0	10.0
1	RF (SMOTE)	56.0	1.0

(i) Velocity

Rank	Treatment	Med	IQR
1	CART (SMOTE)	63.0	1.0
2	RF (SMOTE)	68.0	2.0
3	CART	70.0	2.0
3	RF	70.0	2.0

(j) Xalan

Rank	Treatment	Med	IQR
1	RF	24.0	1.0
2	CART	52.0	18.0
2	CART (SMOTE)	59.0	2.0
2	RF (SMOTE)	60.0	1.0