

READING ASSIGNMENT V

LOOKING FOR BUGS IN ALL THE RIGHT PLACES

By - Robert M. Bell, Thomas J. Ostrand, Elaine J. Weyuker

II. IMPORTANT KEYWORDS

ii1. Regression Analysis

Statistical method for determining relationship between a known and an unknown variable given a population size.

ii2. Misclassification Rates

This would be considered as ratio of data that has been classified wrongly with respect to the entire data set.

ii3. Over dispersion

Negative binomial models allows for some degree of additional variability in faults. This is studied in the paper under the variable over dispersion

ii4. Software Testing

Investigation to provide or judge information about the quality of the software.

III. BRIEF NOTES

iii1. Motivation

With size of software systems increasing, the number of bugs are also going to increase. This paper investigates use of binomial regression model on large industrial systems to predict which file is most likely going to have a bug. This can be used later to localize bugs in software system and concentrate software testing efforts on that system.

iii2. Related Work

The authors have done extensive work in the predicting buggy software files. In the paper "Where the bugs are." [1] published in 2004, they had concluded that the binomial search method is one of the most accurate methods for software bug prediction. Graves et al. [2] have performed study to find out the characteristics of modules associated with faults. Denaro et al. have used logistic regression to construct similar model.

iii3. Hypothesis

Based on accuracy this model can be used for finding files in the software systems which would have higher probability of having bugs. This might help concentrate testing efforts. However, with new study coming as of today, the accuracy of this

model is left wanting. There can be improvement needed in selecting the query and type of data set.

iii4.Results

The results have indicated the below:-

- If predictor variables are fixed, the co-efficient directly relates predictor variables and faults associated with it.*
- When studied over a period of time, with age of the file the number of faults associated with it gradually decreased.*
- For inventory systems, authors concluded that using square root of number of faults in prior releases as a predictor variable has increases the accuracy of the model.*

IV. IMPROVEMENTS

iv1. There model does not penalize for predicting wrong files. Having a variable to penalize wrong listing of files can improve accuracy.

iv2. Software systems typically have bugs in files that have been recently changed, the authors can use a bump up variable for those files.

Relation to Original Paper

The original paper uses the results in this paper and then tries to use Topic-Based Search to bug localization.

References:-

- 1. Where the Bugs are - TJ Ostrand and EJ Weyunkar, 2002*
- 2. Fault incidence using software change history - JTL Graves, AF Karr, JS Marron, H. Siy, 2002*