Improving the Accuracy of Space Mission Software Anomaly Frequency Estimates

Allen P. Nikora, *Member, IEEE,* and Galen Balcom

Abstract—**Anomaly data can be used to estimate baseline values for operational mission software anomaly frequencies; these estimates can be used for future missions to determine whether software reliability is improving. The accuracy of anomaly frequency estimates can be affected by characteristics of the anomaly data and the problem reporting system maintaining that data. We have been using text mining and machine learning techniques to address one of these issues, in which the number of software-related anomalies is incorrectly reported because the problem reporting system does not tag them correctly. Results to date indicate that these techniques may substantially increase the accuracy of anomaly frequency estimates.**

Index Terms—**error estimation, failure analysis, software reliability, text processing.**

I. Introduction

We have been working since 2006 to develop a baseline value for the frequency with which anomalies have been observed for current and historic JPL space mission on-board and ground-based software during mission operations. An accurate estimate would serve as a baseline against which the reliability of future space mission software could be evaluated. Based on an analysis of nearly 8000 anomaly reports for 11 missions, we developed an initial estimate of anomaly frequency and performed simple trend analyses to determine whether that frequency was changing from mission to mission. Our findings were that the mean frequencies at which anomalies were observed was $O(10^{-4})$ anomalies/hour for on-board software and $O(10^{-3})$ for ground-based software. Based on the available data, the anomaly frequency appeared to be increasing from mission to mission at a super-linear rate, for both on-board and ground-based software.

Several characteristics of the anomaly data analyzed for the earlier study had the potential to affect the validity of its results. One of these was the possibility that the number of software-related anomalies may not be accurately documented in the institutional anomaly reporting and tracking system. For this paper, we focused on the case in which anomalies may be incorrectly labeled as to cause: each anomaly tracked in the reporting system has a field which is supposed to label the anomaly by type (e.g., hardware, software, procedural error), but the value entered into that field may not be accurate. A detailed analysis of a sub-set of the anomalies for one of the projects analyzed in the earlier study found that nearly all of the anomalies reported as being software-related appeared to be correctly labeled. However, among the anomalies labeled as non-software, detailed analysis of the problem description, verification, and final corrective action text indicated that a significant number of these anomalies were actually software-related, enough to double the number of anomalies initially identified as software.

The number of software anomalies may be overcounted in the anomaly reporting system. Discussions with a flight software development team member for one of the missions analyzed in the earlier study indicated that the default value of the field identifying the cause of the anomaly was either "flight software" or "ground software", and that these values may not have been changed even if the final determination was that the anomaly was not caused by software.

We applied text mining and machine learning techniques as implemented in the WEKA data mining tool to address this issue. Using a training set developed from a detailed analysis of approximately 800 anomalies recorded during planetary mission operations, we were able to develop learning models capable of distinguishing between software and non-software related anomalies with detection rates of 80% or more, and false positive rates of 20% of less. We are currently applying the classification models learned from the training set to the remainder of the anomalies analyzed for the earlier study, and comparing the performance of these models on the training data to their performance on the remainder of the anomaly reports. If the performance of the models on remainder of the anomaly reports is similar to their performance on the training data, the accuracy of our estimates of software anomaly frequencies and anomaly frequency trends will be substantially improved. We are also investigating the use of text mining and machine learning to discriminate between different types of software anomalies, with the goal of improving our understanding of what types of software anomalies are most frequently encountered, and where they originate.

Manuscript received December 1, 2008. The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored as part of the Software Assurance Research and Ultra-Reliability Programs, funded by the National Aeronautics and Space Administration's Office of Safety and Mission Assurance.

A. P. Nikora is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA (phone: 818-393-1104; fax: 818-393-1362; e-mail: Allen.P.Nikora@jpl.nasa.gov).

G. Balcom is with California State University, Fresno, CA 93740 USA. (e-mail: gbalcom@csufresno.edu).