Globus An International Journal of Management & IT
A Refereed Research Journal
Vol 7 / No 2 / Jan-Jun 2016 ISSN: 0975-721X

A STUDY ON SOFTWARE TESTING WITH DATA MINING

*Neeta

Abstract

Software testing activities are usually planned by human experts, while test automation tools are limited to execution of pre-planned tests only. Regression analysis can imply a far wider range of statistical procedures than often appreciated. In a number of common data mining procedures are discussed within a regression framework. The author discussed one of the newly emerging challenges is how to minimize the amount of retesting after the software is modified. Some people believe that the program correctness can never be demonstrated through software testing.

Data mining algorithms need a technique that partitions the domain values of an attribute in a restricted set of ranges, only because considering every possible ranges of domain values is infeasible. Genetic Algorithm (GA) is a self-adaptive optimization searching algorithm.

Keywords: *Regression analysis, data mining, domain values, Genetic Algorithm*

Introduction

Data, coupled with new data analytics, challenges established epistemologies across the sciences, social sciences and humanities, and assesses the extent to which they are engendering paradigm shifts across multiple disciplines. Software testing activities are usually planned by human experts, while test automation tools are limited to execution of pre-planned tests only. Regression analysis can imply a far wider range of statistical procedures than often appreciated. In a number of common data mining procedures are discussed within a regression framework. In particular, it critically explores new forms of empiricism that declare 'the end of theory', the creation of data-driven rather than knowledge-driven science, and the development of digital humanities and computational social sciences that propose radically different ways to make sense of culture, history, economy and society. It is argued that: (1) Data and new data analytics are disruptive innovations which are reconfiguring in many instances how research is conducted; and (2) there is an urgent need for wider critical reflection within the academy on the epistemological implications of the unfolding data revolution, a task that has barely begun to be tackled despite the rapid changes in research practices presently taking place. After critically reviewing emerging epistemological positions, it is contended that a potentially fruitful approach would be the development of a situated, reflexive and contextually nuanced epistemology.

Evaluation of test outcomes is also associated with a considerable effort by software testers who may have imperfect knowledge of the requirements specification. Not surprisingly, this manual approach to software testing results in heavy losses to the world's economy. Data mining algorithms can be efficiently used for automated modeling of tested systems. Induced data mining models can be utilized for recovering system requirements, identifying equivalence classes in system inputs, designing a minimal set of regression tests, and evaluating the correctness of software outputs.

These include non-parametric smoothers, classification and regression trees, bagging, and random forests. In each case, the goal is to characterize one or more of the distributional features of a response conditional on a set of predictors.

**Research Scholar, Swami Vivekanand University, Sagar, MP*

The most important factor to evaluate the software performance is how well it works and it becomes an essential activity in software engineering field these days. Testing is widely used in industry for quality assurance. Indeed, with the complexity, variety of software growing continuously, ensuring that it behaves according to the desired levels of quality and dependability becomes more crucial, and increasingly difficult and expensive. One of the important and crucial tasks in software engineering process is estimating the cost of the future software and tries to make it as realistic as possible and reduce the cost of development as well. Each and every research on software testing start from claiming that testing is a very expensive activity. A recent study by the National Institute of Standards & Technology found that "the national annual cost of inadequate infrastructure for software testing is estimated to range from $22.2 to $59.5 billion". One of the newly emerging challenges is how to minimize the amount of retesting after the software is modified. Some people believe that the program correctness can never be demonstrated through software testing.

Evaluation of test outputs is also associated with a considerable effort by human testers who often have imperfect knowledge of the requirements specification. Not surprisingly, this manual approach to software testing results in heavy losses to the world's economy. The costs of the so-called "catastrophic" software failures (such as Mars Polar Lander shutdown in 1999) are even hard to measure. In this paper, we demonstrate the potential use of data mining algorithms for automated induction of functional requirements from execution data. The induced data mining models of tested software can be utilized for recovering missing and incomplete specifications, designing a minimal set of regression tests, and evaluating the correctness of software outputs when testing new, potentially flawed releases of the system.

Review of Literature

J. Whittaker et.al (2010) The process of software testing includes four phases namely modeling the software environment, selection of test cases, running and evaluating test cases and measuring testing progress.

M. Vanmali, M. Last and A. Kandel et.al (2012) Traditionally this step was done manually by the human tester, which required a lot of time. As the software systems are growing larger the burden on the human tester is increasing. Using an automated oracle to support the activities of human tester can reduce the cost of the testing process and hence the related maintenance costs.

Dave Kelly et.al (2010) Software testing using an automatic test program will generally avoid the errors that humans make when they get tired after multiple repetitions. The test program won't skip any tests by mistake. The test program can also record the results of the test accurately.

According to Pfleeger et.al (2011) A program fails when it does not do what it is required to do .The purpose of testing a program is to discover faults that cause the system to fail rather than proving the program correctness.

Kaner, C., Falk, J. et.al (2009) A successful test should reveal a problem in software; tests that do not expose any faults are useless, since they hardly provide any indication that the program works properly.

El-Ramly, M. et.al (2012) The activities of system testing include function testing, performance testing, acceptance testing, and installation testing Ideally, a minimal test suite can be generated from a complete and up-to-date specification of functional requirements. Unfortunately, frequent changes make the original requirements documentation, even if once complete and accurate, hardly relevant to the new versions of software.

Software Testing

Testing is indispensible phase of software development life cycle. Debugging is the search for cause of defects. Testing leads to uncovering problems which enhances further debugging. Software deployed without testing leads to unreliability. Hence testing the software to the full extend is a necessary task while building a software. Testing the software implies executing possible test cases. The extend of testing can be evaluated using several techniques like path coverage, conditional coverage, code coverage etc
.
This phase of Software Development Life Cycle is the most expensive phase. It requires lot of time and effort. Hence optimization of test cases is a must. But first we need to see how a test case look like.

A test case is a collection of different attributes of the software. Attributes are the inputs to the software. So a test case can be compared to a touple in a database table. It has ID, attribute1, attribute2, ……..attribute n. A good test case is one that is able to find faults with the software. Hence the output of test case is a pass/fail. A test suite is a collection of automated generated test cases for a particular software. But due to the process of automation redundancy can be initiated

in the process of test data generation. Redundancy is the repetition of data, between one test case and the other. So optimization of test suite is important to achieve by which lot of time can be saved from executing redundant or unnecessary test cases. The behavioral patterns exhibited by the test suite helps us in this process of automation.

Due to the development in software processes, a lot of automation work is carried out in all of its activities. Like so automation is carried out in generating test cases also which enhanced the functionalities in testing phase. An automated test data generator is a program built with the feature of generating inputs to the software by considering business rules and input domain.

In spite of a lot of care taken in generating test cases significant amount of data is replicated in different test cases. This replicated data isn't visible enough to capture unless and until we use sophisticated techniques like data mining.

Data Mining

Data mining is a semi automated process of finding patterns in the data. It is basically knowledge discovery in data. This knowledge discovered can be represented by a set of rules, equations relating different variables and other mechanisms of predicting outcomes.

The manual component of data mining is the preprocessing phase where data is prepared acceptable by the algorithms and post processing phase involving discovering patterns to find out new ones that are useful.

There are three main techniques in data mining classification, association rules and clustering. Classification is a technique that classifies data into different classes by building models like decision trees. By using these models it predicts the behavior of future data. Association rules are the techniques used to find relationships or associations between different entities of an instance. With these associations we can predict the nature of one when the other changes.

Clustering is a technique used in finding clusters of points in the given data. In other words clustering is grouping together similar points into a single cluster. This behavior of grouping can be found out by different metrics like distance, density and grid based approaches. Within a cluster all set of points in that cluster are found to have similar behavior. In order to ease this process of data clustering, in the next section we introduce a tool called weka which helps us in filling the gap between the process of software testing and knowledge mining.

Data mining has emerged as one of the major research domain in the recent decades in order to extract implicit and useful knowledge. This knowledge can be comprehended by humans easily. Initially, this knowledge extraction was computed and evaluated manually using statistical techniques. Subsequently, semi-automated data mining techniques emerged because of the advancement in the technology. Such advancement was also in the form of storage which increases the demands of analysis. In such case, semi-automated techniques have become inefficient. Therefore, automated data mining techniques were introduced to synthesis knowledge efficiently. Consequently, in this paper, we focused on automated data mining techniques. We have decided to critically reviewed literature on these techniques. We have highlighted the strengths and limitations of these automated techniques. Consequently, the significance of this paper is to provide a useful resource to academia as well as researchers in the form of concise literature on automated data mining techniques.

Conclusion

Software testing is used to identify the defects, and improve the quality of software cost. Automation Testing tools are outcome is associated with the effort by software testers. We used to evaluate the fitness of Genetic algorithm for selecting the best possible Test method. This integration will help to improve the common performance of genetic algorithm. To improve testing productivity and reduce the costs, it is highly desirable to automate test generation and execution. The extensive software testing is feasible so it becomes difficult for even medium sized software. To identify the testing efficiency may be increased by using the Keyword- Automated Testing, Data Mining, Genetic Algorithm, Selenium, Load Runner IDE. Software Testing is widely used in industry for quality assurance: indeed, by directly scrutinizing the software in execution, it provides a realistic feedback of its behavior and as such it remains the inescapable complement to other analysis techniques. Software testing is approved as confirmation and validation process that a computer program or application or product should meet the prerequisite that is used in software designing and development.

The design of the data mining based on the background for automated black-box testing. The Random Tests Generator (RTG) Module obtains the list of system inputs, their types (discrete, continuous, etc.), and ranges from the Specification of System Inputs (SSI). Data mining is the process of extracting knowledge hidden from large databases. The knowledge must be new and one

must be proficient to utilize it. Information discovery differs from conventional data retrieval from databases.

Conventional - Traditional DBMS, database data retrieved is not clear. The procedure to discover the Data mining finds these patterns and relationships using data analysis tools and techniques to construct models. To describe this (i) predictive model, which utilizes information with identified results to develop a model that can be used explicitly to expect values. Another is descriptive model, which describes patterns in accessible information.

It is a dominant new knowledge with grand potential to help companies focus on the most important information in their data warehouses. Data mining tools forecast potential trends and behaviors, allow businesses to make practical, knowledge-driven decisions. Data mining has been applied effectively in business environment and also in other fields such as weather conditions forecast, medicine, healthcare, insurance, transportation, government and etc. Data mining brings a set of advantages when used in a specific industry. The main purpose of data mining is to extract patterns from the information at hand, increase its fundamental value and transfer the data into knowledge. Genetic algorithms discover application in bioinformatics, phylogenetic, computational science, engineering, economics, chemistry, manufacturing, mathematics, physics and other fields. A. Use of Genetic Algorithm in Data Mining In this paper, we discuss the applicability of a genetic based algorithm to the search method in data mining. Data mining algorithms need a technique that partitions the domain values of an attribute in a restricted set of ranges, only because considering every possible ranges of domain values is infeasible. Genetic Algorithm (GA) is a self-adaptive optimization searching algorithm.

References

1. Whittaker J., 2010, What Is Software Testing? And Why Is It So Hard?, IEEE Software, Vol 17, No 1,.
2. Vanmali M., Last M. and Kandel A., 2012, Using a Neural Network in Software Testing \ Process, International Journal of Intelligent Systems, Vol 17, No 1, pp. 45-62,.
3. Dave Kelly, 2010 Software Test Automation and the Product Life Cycle", Symantec Corporation, Vol 13, Issue 10.
4. Pfleeger S.L., 2011 Software Engineering: Theory and Practice, 2nd Edition, Prentice-Hall.
5. Kaner C., Falk J. and Nguyen H.Q., 2009, Testing Computer Software, Wiley.
6. Ramly El, Stroulia M. & Sorenson E., 2012 From Run-time Behavior to Usage Scenarios: An Interaction-pattern Mining Approach, Proceedings of KDD- Edmonton, Canada, ACM Press, pp. 315 - 327.