

Chapter 85

Investigation of Software Reliability Prediction Using Statistical and Machine Learning Methods

Pradeep Kumar

Maulana Azad National Urdu University, India

Abdul Wahid

Maulana Azad National Urdu University, India

ABSTRACT

Software reliability is a statistical measure of how well software operates with respect to its requirements. There are two related software engineering research issues about reliability requirements. The first issue is achieving the necessary reliability, i.e., choosing and employing appropriate software engineering techniques in system design and implementation. The second issue is the assessment of reliability as a method of assurance that precedes system deployment. In past few years, various software reliability models have been introduced. These models have been developed in response to the need of software engineers, system engineers and managers to quantify the concept of software reliability. This chapter investigates performance of some classical and intelligent machine learning techniques such as Linear regression (LR), Radial basis function network (RBFN), Generalized regression neural network (GRNN), Support vector machine (SVM), to predict software reliability. The effectiveness of LR and machine learning methods is demonstrated with the help of sixteen datasets taken from Data & Analysis Centre for Software (DACS). Two performance measures, root mean squared error (RMSE) and mean absolute percentage error (MAPE) is compared quantitatively obtained from rigorous experiments.

INTRODUCTION

Software reliability modeling has gained a lot of importance in many critical and daily life applications, which has led to the tremendous work being carried out in software reliability engineering. Software reliability growth models (SRGMs) successfully have been used for estimation and prediction of the number of errors remaining in the software. The software practitioners and potential users can assess the current and future reliability through testing using these *SRGMs*. Many analytical models such as times-between-failures model, nonhomogeneous Poisson process (NHPP) model, Markov processes and operational profile model has been proposed in past four decades for software reliability prediction. The two broad categories of SRGMs include parametric models and non-parametric models. Most of the parametric SRGM models are based on NHPP which has been widely used successfully in practice. The non-parametric SRGM models based on machine learning are more flexible which can predict reliability metrics such as cumulative failures detected, failure rate, time between failures, next time to failures. Both parametric and non-parametric models can be used to estimate the current reliability measures and predict their future trends. Therefore, SRGMs can be used as mathematical tools for measuring, assessing and predicting software reliability quantitatively.

Despite the application of various machine learning methods in past few decades, non-homogeneous Poisson process (NHPP) based models has remained one of the most attractive reliability growth models in monitoring and tracking reliability improvement. However, due to their hard-core assumptions, validity and relevance in the real-world scenario have limited their usefulness. On the other hand, learning and generalization capability of artificial neural networks (ANNs), and its proven successes in complex problem solutions has made it, a viable alternative for predicting software failures in the testing phase. The main advantages of ANNs over NHPP based models is that it requires only failure history as inputs and no assumptions, or a priori postulation of parametric models is required. Several regression techniques such as linear regression and machine learning methods (DTs, ANNs, SVMs, GA) have been proposed in literature for predicting software reliability. The major challenges of these models do not lie in their technical soundness, but their validity and applicability in real world projects in particular to modern computing system. Linear regression (LR) is the most widely used method and easily understood but it hardly works well on real-life data. Since, LR is restricted to fitting straight line functions to data and hence not suited well for modeling non-linear functions. Some empirical studies based on multivariate linear regression and neural network methods have been carried out for prediction of software reliability growth trends.

However, multivariate linear regression method can address linear relationship but require large sample size and more independent variables. The use of support vector machine (SVM) approach in place of classical techniques has shown a remarkable improvement in the prediction of software reliability in the recent years. The design of SVM is based on the extraction of a subset of the training data that serves as support vectors and therefore represents a stable characteristic of the data. GRNN-based reliability prediction model incorporating the test coverage information such as blocks and branches is applied for software reliability prediction. The prediction accuracy of software reliability models can be further improved by adding other important factors affecting the final software quality such as historical information from software development like capability of developers, testing effort and test coverage. SVM represent the state of the art due to their generalization performance, ease of usability and rigorous theoretical foundations that practically can be used for regression solving problems.

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/investigation-of-software-reliability-prediction-using-statistical-and-machine-learning-methods/252104

Related Content

Spatial Modeling and Geovisualization of Rental Prices for Real Estate portals

Harald Schernthanner, Hartmut Asche, Julia Gonschorek and Lasse Scheele (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 962-977).

www.irma-international.org/chapter/spatial-modeling-and-geovisualization-of-rental-prices-for-real-estate-portals/252065

Designing Extreme Learning Machine Network Structure Based on Tolerance Rough Set

Han Ke (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 263-282).

www.irma-international.org/chapter/designing-extreme-learning-machine-network-structure-based-on-tolerance-rough-set/252029

A Novel Anti-Obfuscation Model for Detecting Malicious Code

Yuehan Wang, Tong Li, Yongquan Cai, Zhenhu Ning, Fei Xue and Di Jiao (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1556-1576).

www.irma-international.org/chapter/a-novel-anti-obfuscation-model-for-detecting-malicious-code/252098

A Hybrid GSA-K-Mean Classifier Algorithm to Predict Diabetes Mellitus

Rojalina Priyadarshini, Rabindra Kumar Barik, Nilamadhab Dash, Brojo Kishore Mishra and Rachita Misra (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 589-603).

www.irma-international.org/chapter/a-hybrid-gsa-k-mean-classifier-algorithm-to-predict-diabetes-mellitus/252045

Diagnosis Rule Extraction from Patient Data for Chronic Kidney Disease Using Machine Learning

Alexander Arman Serpen (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1165-1174).

www.irma-international.org/chapter/diagnosis-rule-extraction-from-patient-data-for-chronic-kidney-disease-using-machine-learning/252076