Software Reliability Prediction Using Cuckoo Search Optimization, Empirical Mode Decomposition, and ARIMA Model: CS-EEMD-ARIMA Based SRGM

Ankur Choudhary, Gautam Buddha University, Department of Computer Science and Engineering, Noida, India
Anurag Singh Baghel, Gautam Buddha University, Department of Computer Science and Engineering, Noida, India

ABSTRACT
The demand of highly reliable and superior quality open source software’s are increasing day by day. This demand force software developers to improvise the reliability of their software’s. Authors have proposed several parametric and non-parametric software reliability models but they have their own limitations, like parametric model suffer from unrealistic model assumptions, operating environment condition dependencies. In contrast to parametric, non-parametric models overcome these issues but they are computationally costlier. So, the scope of optimization or development of new reliable model still exists. This paper presents an effective software reliability modeling based on Cuckoo Search optimization, Ensemble Empirical Mode Decomposition and ARIMA modeling of time series to provide more accurate prediction. Extensive experiments on 5 real datasets is conducted and results are collected. The analysis of results indicates the superiority of proposed technique over existing parametric and non-parametric models for open source software’s and propriety software’s.

KEYWORDS
ARIMA, Cuckoo Search Optimization, EEMD, Software Reliability Growth Model, Time Series Model

DOI: 10.4018/IJOSSP.2016100103

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

Open source and propriety software’s is competitor on quality and reliability metrics. Software reliability is an important software quality metric, which describes the failure free functioning of software underneath specified conditions and time period. It also develops trust among the customers and helps the software designers to manage the upcoming requirements and maintenance activities for high-quality software development.

Literature states that the software reliability relies on various factors during software development life cycle (SDLC) such as (Kapoor, Pham, Gupta, & Jha, 2011):

1. The imperfect requirement gathering may affect the software design, as well as can be a cause of failure.
2. The design and implementation quality of software may affect its reliability.
3. The imperfect testing of software increases the chance of failure occurrence.
4. The incomplete validation may decrease the reliability of software.

This paper mainly focuses on the models based on software testing data, known as software reliability growth model (SRGM). These models work on the relationship of software failures with respect to time (Amin, Grunse, & Colman, 2013; Choudhary, Baghel, & Sangwan, 2017). The mathematical modeling theory classifies SRGMs into two categories (Li, Lu, Zeng, & Huang, 2012):

1. Parametric SRGMs
2. Non-Parametric SRGMs

The parametric SRGMs are most commonly used mathematical models for software reliability prediction. These models make use of several unrealistic assumptions regarding software faults and testing process (Costa, Vergilio, Pozo, & Souza, 2005; Goel, 1985) and are not purely generalized for all types of software (Choudhary et al., 2017; Lo, 2011; Musa, Iannino, & Okumoto, 1987). To overcome such limitations non-parametric or data driven SRGMs have been proposed by researchers. These models make use of machine learning approaches such as: artificial neural networks (ANN), bayesian networks, support vector machine (SVM), (Amin et al., 2013; Bo Yang & Xiang Li, 2007; Jin, 2011; Li, Gu, Feng, & Chen, 2009; Lo, 2011). These models map the stochastic nature of software failures and provide better quality software failure prediction (Karunanithi, Whitley, & Malaiya, 1992; Sitte, 1999). But large training dataset and computational cost per prediction is major drawback for ANN and other machine learning techniques (Amin et al., 2013). Few authors have proposed an effective alternative solution as data driven approaches by treating the failure data as time series. There are several time series modeling based SRGMs are proposed (Amin et al., 2013; Bao, Yi, & Chen, 2012; Lo, 2011). Auto regressive integrated moving average (ARIMA) model is the most explored and reliable time
Higher Education and FOSS for E-Learning: The Role of Organizational Sub-Cultures in Enterprise-Wide Adoption
www.igi-global.com/chapter/higher-education-foss-learning/46307?camid=4v1a

A Model for the Successful Migration to Desktop OSS
www.igi-global.com/chapter/model-successful-migration-desktop-oss/21186?camid=4v1a