Journal of Information Technology Research
Volume 13 • Issue 1 • January-March 2020

Software Aging Forecast Using Recurrent SOM with Local Model

Yongquan Yan, School of Statistics, Shanxi University of Finance and Economics,, Taiyuan, China

ABSTRACT

Studies of software aging problems are important since they are related to QoS. Previous studies have used many methods to guarantee QoS. In this article, a recurrent self-organizing map with multi-layerperceptron is proposed to forecast resource consumption in a web server which suffered from a software aging problem. First, a resource consumption series in a web server is split into p dimensional space vectors. Second, the split series is clustered into local sets by using a recurrent self-organizing map. Last, a local prediction method called multi-layerperceptron is used to predict on each local set. The results indicated that the recurrent self-organizing map with multi-layerperceptron generates a slightly better estimation than multi-layerperceptron and autoregressive integrated moving average in the resource consumption predictions of system and application level of web server.

KEYWORDS

MLP, Qos, RSOM, Software Aging, Web Server

INTRODUCTION

As increasing software complexity, some phenomena, such as performance degradation, unplanned downtime, have been found in many systems: operating system (Cotroneo, 2010), web server (Grottke, 2006), communication system (Hoffmann, 2007), android system (Huo, 2018), stream processing system (Ficco, 2018), and so on. These phenomena are called software aging problems (Varasteh, 2017), which are influenced by many factors, such as memory leak, unreleased file connection, round-off error due to software bugs. Software bugs can be classified into two types of bugs, Mandelbug (Grottke, 2007) and Bohrbug. Software aging bug is mainly involved in Mandelbug which can be hardly reproduced in the same situations. So, the traditional tolerance technologies, such as multiple replicas, cannot solve the software aging problems. These problems can be settled by a method called software rejuvenation, which reboot the software system suffered from software aging problems, clean the internal states and make the software system enter into a fresh state.

The main motivation for software aging and rejuvenation research is the desire to find software aging phenomenon in advance and make a software system enter into a robust state with a high service quality. Since resource consumption is a key factor to influence software state and software state can be represented as time series, many studies concentrate on building a series processing model to forecast the future events of software state.

A resource consumption series is composed of a natural measurement or observation process which is made sequentially in time. All kinds of methods, which can be classified into linear methods and nonlinear methods, are used to forecast resource consumption. Among various linear methods, autoregressive and autoregressive integrated moving average (in abbreviated form, ARIMA) (Box,

DOI: 10.4018/JITR.2020010103

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

1970) models are the most used approaches in practice. Although a resource consumption series owns linear characteristics in a short time span, it shows nonlinear feature in the long run. Artificial neural network methods, which are biologically inspired models that are typically made up of simple connected computing units in some ways, are most used nonlinear methods.

Also, linear and nonlinear models (Zhang, 2018) can be divided into global and local models by the modeling process. In global models, only a model is used to train the collected data, such as autoregressive model and radial basis function network. Global methods give the best results with stationary resource consumption series. However, when the resource consumption series is nonstationary, it is difficult to find a suitable model and often some methods, such as difference method, can be used to make the series stationary.

In the past years, local methods, such as self-organizing map, have arisen enough interest, since they can solve some problems in the global methods (Singer, 1992). In the local methods, the dataset is divided into some smaller datasets and these datasets are executed with some clustering or quantization algorithms such as k- means or neural gas. After clustering the data, the local method is used to train the local data.

In order to train the model, a window method is used to divide the resource consumption series into input vectors which are used as input variables. Thus, the length of the window influences the model prediction precision. When the resource consumption series is stationary with certain time lag, dividing the data with a window cannot lose too much information. When the series is nonstationary, choosing a proper window length is very difficult and this is a considerable problem for global models.

In this work, we propose a new method based on recurrent self-organizing map (RSOM) architecture and use it to improve the precisions of resource consumption suffered from software aging problems. RSOM is applied to cluster resource consumption data to the local datasets and a nonlinear method is used to fit the data based on the local datasets. Compared to the other methods, the proposed method in this work can trace the past input values with a recurrent network, which means that the temporal context can be saved and used in the training procedure. The rest of the paper is organized as following. In section two, related work is discussed. In section three, the new method is proposed and explained. In section four, the experiment is introduced and the prediction results of the proposed method are compared with linear and nonlinear methods. In the last, conclusions and discussions are made.

RELATED WORK

In order to find software aging in advance, the measurement-based approaches are used to determine which parts of software systems are in aging phenomena based on periodically collecting different types of software system metrics such as memory usage, CPU consumption, response time. Time series analysis methods for software aging problems are used to: (1) detect the presence of a degradation trend; (2) predict the key software system metrics to find software aging ahead of time. Some trend detection algorithms, such as Mann-Kendall (Garg, 1968) and Seasonal Kendall test (Sen, 1968; Theil, 1992) were used to detect software aging emergence, when hypothesis tests can be accepted. Since any distribution assumptions on the data are not made, these methods are non-parametric tests and more robust to outliers compared to parametric tests. However, Mann-Kendall often combined with Seasonal Sen's trend estimation algorithm suffered from some false alarms (Zheng, 2014; Machida, 2013) and was suitable for linear trends only, sensitivity to noise and high computational complexity.

So, some time-series and machine learning algorithms are used to forecast resource consumption, which is a key element for software aging occurrence. Li et al. (2002) used ARMA models to predict resource consumption in Apache web server. Magalhaes et al. (2010) proposed a framework to detect abnormal phenomena by using ARIMA and Hot-Winters. At run-time, all collected running parameters are trained and predicted by ARIMA and Hot-Winters. However, these methods have an assumption that time series data are linear, which can hardly find in real situations. Machine learning approaches

12 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/article/software-aging-forecast-using-recurrent-som-with-local-model/240720

Related Content

Heuristics in Medical Data Mining
Susan E. George (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 1723-1726).*
www.irma-international.org/chapter/heuristics-medical-data-mining/13808

Computer Service Support at Glenview Hospital
Martha Garcia-Murillo, Paula Maxwell, Simon Boyce, Raymond St. Denisand William Bistline (2003). *Annals of Cases on Information Technology: Volume 5 (pp. 387-400).*
www.irma-international.org/article/computer-service-support-glenview-hospital/44554

Reference Model Management
Oliver Thomas (2009). *Selected Readings on Information Technology Management: Contemporary Issues (pp. 1-20).*
www.irma-international.org/chapter/reference-model-management/28658

Information and Knowledge Perspectives in Systems Engineering and Management for Innovation and Productivity Through Enterprise Resource Planning
Stephen V. Stephensonand Andrew P. Sage (2007). *Information Resources Management Journal (pp. 44-73).*
www.irma-international.org/article/information-knowledge-perspectives-systems-engineering/1311

Service Description Ontologies
Julia Kantorovitchand Eila Niemelä (2009). *Encyclopedia of Information Science and Technology, Second Edition (pp. 3445-3451).*
www.irma-international.org/chapter/service-description-ontologies/14085