1. Predicting Crash Fault Residence via Simplified Deep Forest Based on A Reduced Feature Set

Zhao, K. (1); Liu, J. (1); Xu, Z. (2); Li, L. (3); Yan, M. (2); Yu, J. (1); Zhou, Y. (4)
Source: *2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC)*, p 242-52, 2021;
ISBN-13: 978-1-6654-1403-6; **DOI:** 10.1109/ICPC52881.2021.00031; **Conference:** 2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC), 20-21 May 2021, Virtual Conference, Spain; **Publisher:** IEEE Computer Society, Los Alamitos, CA, USA
Author affiliation: (1) Wuhan University, School of Computer Science, China (2) Chongqing University, School of Big Data and Software Engineering, China (3) Monash University, Faculty of Information Technology, Melbourne, VIC, Australia (4) Syracuse University, College of Engineering and Computer Science, Syracuse, NY, United States
Abstract: The software inevitably encounters the crash, which will take developers a large amount of effort to find the fault causing the crash (short for crashing fault). Developing automatic methods to identify the residence of the crashing fault is a crucial activity for software quality assurance. Researchers have proposed methods to predict whether the crashing fault resides in the stack trace based on the features collected from the stack trace and faulty code, aiming at saving the debugging effort for developers. However, previous work usually neglected the feature preprocessing operation towards the crash data and only used traditional classification models. In this paper, we propose a novel crashing fault residence prediction framework, called ConDF, which consists of a consistency based feature subset selection method and a state-of-the-art deep forest model. More specifically, first, the feature selection method is used to obtain an optimal feature subset and reduce the feature dimension by reserving the representative features. Then, a simplified deep forest model is employed to build the classification model on the reduced feature set. The experiments on seven open source software projects show that our ConDF method performs significantly better than 17 baseline methods on three performance indicators. (0 refs)
Inspec controlled terms: feature selection - pattern classification - program debugging - public domain software - software quality
Uncontrolled terms: crash fault residence prediction - ConDF method - simplified deep forest model - representative features - feature dimension - optimal feature subset - state-of-the-art deep forest model - consistency based feature subset selection method - novel crashing fault residence prediction framework - crash data - feature preprocessing operation - stack trace - crashing fault resides - software quality assurance - reduced feature set
Classification Code: C6110B Software engineering techniques - C6130 Data handling techniques - C6150G Diagnostic, testing, debugging and evaluating systems - C1160 Combinatorial mathematics
IPC Code: G06F7/00 - G06F9/44 - G06F11/36
Treatment: Practical (PRA)
Database: Inspec
Data Provider: Engineering Village
Copyright 2021, The Institution of Engineering and Technology