

Faster Detecting of Duplicate Bug Report with Maintaining Accuracy

Zahra Aminoroaya¹

Behzad Soleimani Neysiani²

Mohammad Hossein Nadimi Shahraki³

¹Allame Naeini Higher Education Institute, Naein, Isfahan, Iran

²Faculty of Electrical and Computer Engineering, University of Kashan, Kashan, Isfahan, Iran

³Faculty of Computer Engineering, Najafabad Branch Islamic Azad University, Najafabad, Isfahan, Iran

ABSTRACT

Nowadays, duplicate bug reports detection is one of major problems for user bug reports tracking systems. Many researchers used information retrieval tools and methods for solving this problem which are used in this study with new feature extracted based on minimum, maximum and average similar frequent terms in two bug reports. First, 162 features are extracted in 4 large data set repositories as Android, Mozilla, Open Office and Eclipse which contains of new and introduced features in the state of the art. Then dimension reduction methods are used to eliminate negligible importance features and improve the running time of train and testing classification algorithms. Experimental results show the runtime time of the classification algorithms by reduced features reduced from minutes to seconds in comparing to all features, while the accuracy of duplicate detection improved between 1% to 6%. Due to new proposed features, the accuracy and recall of duplicate detection are more than 96% and 90% respectively.

Keywords: Feature Extraction, Duplicate Bug Reports Detection, Classification Algorithms, Accuracy, Run-Time.