

Computer Science, published by China Computer Federation (CCF)  
To appear in 2016/05

## Extraction Approach for Software Bug Report

LIN Tao<sup>1</sup>, GAO Jian-hua<sup>1</sup>, FU Xue<sup>1</sup>, MA Yan<sup>1</sup>, LIN Yan<sup>2</sup>

(Department of Computer Science and Technology, Shanghai Normal University, Shanghai, 200234, China)<sup>1</sup>

(Department of Information Systems, The University of Auckland, Auckland, 92019, New Zealand)<sup>2</sup>

**Abstract:** There are increasingly bug reports in software engineering and developers bewildered by the rapid reports accumulation. Therefore, it is necessary to extract bug report for the task of bug fixing and software reuse, etc. The paper proposes a novel extraction approach. Synonyms were merged into one specific word firstly in the approach. Then, it sets up a vector space model, and some text mining methods, such as TF-IDF and information gain, are presented to collect word features for bug reports specifically, but also there is an algorithm for determining sentence complexity, so as to choose the sentence in long length. This work introduced Bayes classifier into bug report extraction. TPR is increased and FPR is decreased in this approach. The experiment proves that bug report extraction by using data mining and Bayes classifier is competitive through the evaluation of AUC(0.71), F-score (0.80) and Kappa value(0.75).

**Keywords:** Bug report management, Text mining, Bayes classifier, Bug feature, Vector space model, Sentence complexity