

Reading Assignment 9

Bhanu Anand, Esha Sharma, Vinay Todalge

i. Reference to the paper:

Anahita Alipour, Abram Hindle, Eleni Stroulia

[“A contextual approach towards more accurate duplicate bug report detection”](#) Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on 18-19 May 2013.

Relation to First Paper:

This paper refers to our first paper for comparing result. Results from first paper were used as base results against which improvements of this new approach are compared.

ii. Important Key-words:

1. Contextual Information

In addition to IR based textual similarities, authors have used extra information relying on prior knowledge of software quality, software architecture, and system-development topics. These sets of information is called as Contextual information. This information includes Android architecture words, Non- functional requirement words, Android Topic words etc.

2. Textual similarity

Bugs are textual representation of states of machine. While comparing two bugs, textual features such as Description, Version, Summary etc are used. BM25F is such one kind of information retrieval tool used in first paper against which results from this paper are compared.

3. Bug Deduplication

Bug deduplication is process of finding out whether given bug is duplicate of any other bugs present in bug repository.

4. Comparison

While deciding a given bug report is duplicate or not, bug reports are compared to each other using various techniques. Textual based Information Retrieval techniques are mostly used techniques in this area. This paper has suggested to use contextual information as well for comparison.

iii. Brief Note:

1. Motivation:

While textual feature based comparison among bug reports is not accurate enough to identify all duplicates in system, authors had thought of using additional information which helps to get more accuracy in duplicate detection. Additional information is contextual information know in advance such as Architecture words, Requirement terms.

2. Baseline results:

Results from our first paper were used here as baseline results for comparison. Top-k similar reports were retrieved and classification based approach is followed for duplicate detection in first paper with use of BM25F. Only textual features were used for comparing bug reports in first paper.

3. Sampling Procedure:

Each report is preprocessed to remove stop words and stemmed. Preprocessed bug reports with the help of contextual similarity were analyzed for finding out whether given bug is duplicate or not. Textual measures and contextual measure were used before machine learning algorithm is applied.

4. Data:

Author have used bug report from Android System project consisting of 37236 bug reports reported between November 2007 to September 2012.

5. Results:

Results of this paper are compared with results of first paper. Author claims that contextual approach improves accuracy in detecting duplicate bugs up to 11.55%. Reason for selecting firsts paper for comparing results is first paper uses textual similarity very effectively for duplicate detection. Author have wanted to compare results with model which makes use of textual similarity of bug reports very efficiently.

iv. Ways paper would be improved:

1. No standardization in collecting contextual datasets.
2. Automatic bug assignment to respective developer can be done as contextual information present beforehand might easily find respective developer for a corresponding bug.
3. Only bug reports from Android System Project were used for analysis in this paper. While considering contextual information in order to find duplicates, variety number of bugs should be tested. Here, contextual information about only android system is used for finding duplicates.