Bug or not bug? That is the question.

Quentin Perez[*1], Pierre-Antoine Jean[1], Christelle Urtado[1], and Sylvain Vauttier[1]

[1]EuroMov - Digital Health in Motion (Euromov DHM) – IMT - MINES ALES, Institut Mines-Télécom [Paris], Université de Montpellier : $UR_U M_I MT_1 02 -$
$-UniversitédeMontpellierUFRSTAPS700avenueduPicSaintLoup34090Montpellier, France$

Résumé

Nowadays, development teams often rely on tools such as Jira or Bugzilla to manage backlogs of issues to be solved to develop or maintain software. Although they relate to many different concerns (e.g., bug fixing, new feature development, architecture refactoring), few means are proposed to identify and classify these different kinds of issues, except for non mandatory labels that can be manually associated to them. This may lead to a lack of issue classification or to issue misclassification that may impact automatic issue management (planning, assignment) or issue-derived metrics. Automatic issue classification thus is a relevant topic for assisting backlog management. This paper proposes a binary classification solution for discriminating bug from non bug issues. This solution combines natural language processing (TF-IDF) and classification (multi-layer perceptron) techniques, selected after comparing commonly used solutions to classify issues. Moreover, hyper-parameters of the neural network are optimized using a genetic algorithm. The obtained results, as compared to existing works on a commonly used benchmark, show significant improvements on the F1 measure for all datasets.

---

[*]Intervenant