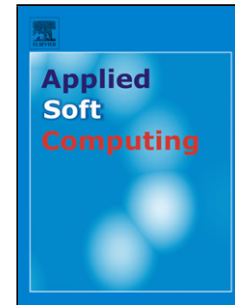


Accepted Manuscript

Title: Learning to classify software defects from crowds: a novel approach

Author: Jerónimo Hernández-González Daniel Rodriguez
Iñaki Inza Rachel Harrison Jose A. Lozano



PII: S1568-4946(17)30661-0
DOI: <https://doi.org/doi:10.1016/j.asoc.2017.10.047>
Reference: ASOC 4542

To appear in: *Applied Soft Computing*

Received date: 1-4-2017
Revised date: 27-10-2017
Accepted date: 31-10-2017

Please cite this article as: Jerónimo Hernández-González, Daniel Rodriguez, Iñaki Inza, Rachel Harrison, Jose A. Lozano, Learning to classify software defects from crowds: a novel approach, *Applied Soft Computing Journal* (2017), <https://doi.org/10.1016/j.asoc.2017.10.047>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Learning to classify software defects from crowds: a novel approach

Jerónimo Hernández-González^{a,*}, Daniel Rodríguez^c, Iñaki Inza^a, Rachel Harrison^d, Jose A. Lozano^{a,b}

^a *Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Donostia, Spain*

^b *Basque Center for Applied Mathematics BCAM, Bilbao, Spain*

^c *Department of Computer Science, University of Alcalá, Madrid, Spain*

^d *Department of Computing, Oxford Brookes University, Oxford, UK*

Abstract

In software engineering, associating each reported defect with a category allows, among many other things, for the appropriate allocation of resources. Although this classification task can be automated using standard machine learning techniques, the categorization of defects for model training requires expert knowledge, which is not always available. To circumvent this dependency, we propose to apply the learning from crowds paradigm, where training categories are obtained from *multiple non-expert* annotators (and so may be incomplete, noisy or erroneous) and, dealing with this subjective class information, classifiers are efficiently learnt. To illustrate our proposal, we present two real applications of the IBM's orthogonal defect classification working on the issue tracking systems from two different real domains. Bayesian network classifiers learnt using two state-of-the-art methodologies from data labeled by a crowd of annotators are used to predict the category (impact) of reported software defects. The considered methodologies show enhanced performance regarding the straightforward solution (majority voting) according to different metrics. This shows the possibilities of using non-expert knowledge aggregation techniques when

*Corresponding author

Email addresses: `jeronimo.hernandez@ehu.eus` (Jerónimo Hernández-González), `daniel.rodriguez@uah.es` (Daniel Rodríguez), `inaki.inza@ehu.eus` (Iñaki Inza), `rachel.harrison@brookes.ac.uk` (Rachel Harrison), `ja.lozano@ehu.eus` (Jose A. Lozano)

Download English Version:

<https://daneshyari.com/en/article/6904285>

Download Persian Version:

<https://daneshyari.com/article/6904285>

[Daneshyari.com](https://daneshyari.com)