Additional Material
Learning to classify software defects from crowds: a novel approach

Jerónimo Hernández-González[a,*], Daniel Rodriguez[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 Alcala, Madrid, Spain*
[d] *Department of Computering, Oxford Brookes University, Oxford, UK*

Abstract

Example of the computation of the formulae of the Section 3 of the paper.

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

1. Estimation of model parameters

Table 1: Example of the annotations of 3 labelers for 4 different examples

$\boldsymbol{l}^1$ ={ Instal., Other, Instal.}	$\boldsymbol{l}^2$ ={ Req., Req., Req.}
$\boldsymbol{l}^3$ ={ Req., Usab., Other}	$\boldsymbol{l}^4$ ={ Usab., Usab., Req.}

Table 2: Example of per-label weights (w_c^a) for the 3 annotators.

$A1 : \{0.7, 0.8, 0.6, 0.7\}$ $A2 : \{0.8, 0.6, 0.7, 0.6\}$ $A3 : \{0.7, 0.7, 0.6, 0.6\}$

Table 3: Computation of Eq. 3 for examples of Table 1 using per-label weights of Table 2

$$F^{\boldsymbol{l}^1}_{Inst} = \frac{1{\cdot}0.7+0{\cdot}0.8+1{\cdot}0.7}{1{\cdot}0.7+0{\cdot}0.8+1{\cdot}0.7+0{\cdot}0.8+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+1{\cdot}0.6+0{\cdot}0.6} = \frac{0.7+0.7}{0.7+0.7+0.6} = 0.7$$

$$F^{\boldsymbol{l}^1}_{Req} = \frac{0{\cdot}0.8+0{\cdot}0.6+0{\cdot}0.7}{1{\cdot}0.7+0{\cdot}0.8+1{\cdot}0.7+0{\cdot}0.8+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+1{\cdot}0.6+0{\cdot}0.6} = \frac{0.0}{0.7+0.7+0.6} = 0.0$$

$$F^{\boldsymbol{l}^1}_{Usab} = \frac{0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6}{1{\cdot}0.7+0{\cdot}0.8+1{\cdot}0.7+0{\cdot}0.8+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+1{\cdot}0.6+0{\cdot}0.6} = \frac{0.0}{0.7+0.7+0.6} = 0.0$$

$$F^{\boldsymbol{l}^1}_{Other} = \frac{0{\cdot}0.7+1{\cdot}0.6+0{\cdot}0.6}{1{\cdot}0.7+0{\cdot}0.8+1{\cdot}0.7+0{\cdot}0.8+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+0{\cdot}0.6+0{\cdot}0.7+1{\cdot}0.6+0{\cdot}0.6} = \frac{0.6}{0.7+0.7+0.6} = 0.3$$

c	$F^{\boldsymbol{l}^2}_c$	$F^{\boldsymbol{l}^3}_c$	$F^{\boldsymbol{l}^4}_c$
$Inst.$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.0}{0.8+0.7+0.6} = 0$	$\frac{0.0}{0.6+0.7+0.7} = 0$
$Req.$	$\frac{0.8+0.6+0.7}{0.8+0.6+0.7} = 1$	$\frac{0.8}{0.8+0.7+0.6} = 0.38$	$\frac{0.7}{0.6+0.7+0.7} = 0.35$
$Usab.$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.7}{0.8+0.7+0.6} = 0.33$	$\frac{0.6+0.7}{0.6+0.7+0.7} = 0.65$
$Other$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.6}{0.8+0.7+0.6} = 0.29$	$\frac{0.0}{0.6+0.7+0.7} = 0$

Table 4: Example of confusion-matrix weights ($W^a_{cc'}$) for the 3 annotators. Note that the diagonals of the matrices are the same per-label weights of Table 2.

$$A1: \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 \\ 0.1 & 0.8 & 0.1 & 0.0 \\ 0.1 & 0.2 & 0.6 & 0.1 \\ 0.2 & 0.0 & 0.1 & 0.7 \end{bmatrix} \quad A2: \begin{bmatrix} 0.8 & 0.1 & 0.0 & 0.1 \\ 0.1 & 0.6 & 0.1 & 0.2 \\ 0.0 & 0.2 & 0.7 & 0.1 \\ 0.2 & 0.0 & 0.2 & 0.6 \end{bmatrix} \quad A3: \begin{bmatrix} 0.7 & 0.2 & 0.1 & 0.0 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.6 & 0.0 \\ 0.1 & 0.2 & 0.1 & 0.6 \end{bmatrix}$$

Table 5: Computation of Eq. 4 for examples of Table 1 using confusion-matrix weights of Table 4

$$F_{Inst}^{\boldsymbol{l}^1} = \frac{1\cdot0.7+1\cdot0.2+1\cdot0.7}{1\cdot0.7+1\cdot0.2+1\cdot0.7+0\cdot0.2+0\cdot0.0+0\cdot0.2+0\cdot0.1+0\cdot0.2+0\cdot0.1+1\cdot0.0+1\cdot0.6+1\cdot0.0} = \frac{0.7+0.2+0.7}{0.7+0.2+0.7+0.6} = 0.73$$

$$F_{Req}^{\boldsymbol{l}^1} = \frac{0\cdot0.2+0\cdot0.0+0\cdot0.2}{1\cdot0.7+1\cdot0.2+1\cdot0.7+0\cdot0.2+0\cdot0.0+0\cdot0.2+0\cdot0.1+0\cdot0.2+0\cdot0.1+1\cdot0.0+1\cdot0.6+1\cdot0.0} = \frac{0.0}{0.7+0.2+0.7+0.6} = 0.0$$

$$F_{Usab}^{\boldsymbol{l}^1} = \frac{0\cdot0.1+0\cdot0.2+0\cdot0.1}{1\cdot0.7+1\cdot0.2+1\cdot0.7+0\cdot0.2+0\cdot0.0+0\cdot0.2+0\cdot0.1+0\cdot0.2+0\cdot0.1+1\cdot0.0+1\cdot0.6+1\cdot0.0} = \frac{0.0}{0.7+0.2+0.7+0.6} = 0.0$$

$$F_{Other}^{\boldsymbol{l}^1} = \frac{1\cdot0.0+1\cdot0.6+1\cdot0.0}{1\cdot0.7+1\cdot0.2+1\cdot0.7+0\cdot0.2+0\cdot0.0+0\cdot0.2+0\cdot0.1+0\cdot0.2+0\cdot0.1+1\cdot0.0+1\cdot0.6+1\cdot0.0} = \frac{0.6}{0.7+0.2+0.7+0.6} = 0.27$$

c	$F_c^{\boldsymbol{l}^2}$	$F_c^{\boldsymbol{l}^3}$	$F_c^{\boldsymbol{l}^4}$
$Inst.$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.0}{0.8+0.2+0.2+0.1+0.7+0.1+0.0+0.1+0.6} = 0$	$\frac{0.0}{0.2+0.2+0.7+0.6+0.7+0.1} = 0$
$Req.$	$\frac{0.8+0.6+0.7}{0.8+0.6+0.7} = 1$	$\frac{0.8+0.2+0.2}{0.8+0.2+0.2+0.1+0.7+0.1+0.0+0.1+0.6} = 0.43$	$\frac{0.2+0.2+0.7}{0.2+0.2+0.7+0.6+0.7+0.1} = 0.44$
$Usab.$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.1+0.7+0.1}{0.8+0.2+0.2+0.1+0.7+0.1+0.0+0.1+0.6} = 0.32$	$\frac{0.6+0.7+0.1}{0.2+0.2+0.7+0.6+0.7+0.1} = 0.56$
$Other$	$\frac{0.0}{0.8+0.6+0.7} = 0$	$\frac{0.0+0.1+0.6}{0.8+0.2+0.2+0.1+0.7+0.1+0.0+0.1+0.6} = 0.25$	$\frac{0.0}{0.2+0.2+0.7+0.6+0.7+0.1} = 0$

2. Estimation of reliability weights for the annotators

Table 6: Computation of per-label weights (w_c^a, Eq. 5) for annotations of Table 1.

c	L_1	L_2	L_3
$Inst.$	$\frac{1\cdot\frac{1}{2}\cdot(0+1)}{1} = 0.5$	0.0	$\frac{1\cdot\frac{1}{2}\cdot(1+0)}{1} = 0.5$
$Req.$	$\frac{1\cdot\frac{1}{2}\cdot(1+1)+1\cdot\frac{1}{2}\cdot(0+0)}{2} = 0.5$	$\frac{1\cdot\frac{1}{2}\cdot(1+1)}{1} = 1.0$	$\frac{1\cdot\frac{1}{2}\cdot(1+1)+1\cdot\frac{1}{2}\cdot(0+0)}{2} = 0.5$
$Usab.$	$\frac{1\cdot\frac{1}{2}\cdot(1+0)}{1} = 0.5$	$\frac{1\cdot\frac{1}{2}\cdot(0+0)+1\cdot\frac{1}{2}\cdot(1+0)}{2} = 0.25$	0.0
$Other$	0.0	$\frac{1\cdot\frac{1}{2}\cdot(0+0)}{1} = 0.0$	$\frac{1\cdot\frac{1}{2}\cdot(0+0)}{1} = 0.0$

Table 7: Computation of confusion-matrix weights ($W^a_{cc'}$, Eq. 6) for annotations of Table 1.

L_1:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{1} = 0.5$	0.0	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{1} = 0.5$
$Req.$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+1) + 1 \cdot \frac{1}{2} \cdot (0+0)}{2} = 0.5$	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{2} = 0.25$	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{2} = 0.25$
$Usab.$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{1} = 0.5$	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{1} = 0.5$	0.0
$Other$	0.0	0.0	0.0	0.0

L_2:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	0.0	0.0	0.0	0.0
$Req.$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+1)}{1} = 1.0$	0.0	0.0
$Usab.$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+0) + 1 \cdot \frac{1}{2} \cdot (0+1)}{2} = 0.5$	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{2} = 0.25$	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{2} = 0.25$
$Other$	$\frac{1 \cdot \frac{1}{2} \cdot (1+1)}{1} = 1.0$	0.0	0.0	0.0

L_3:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{1} = 0.5$	0.0	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{1} = 0.5$
$Req.$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+1) + 1 \cdot \frac{1}{2} \cdot (0+0)}{2} = 0.5$	$\frac{1 \cdot \frac{1}{2} \cdot (1+1)}{2} = 0.5$	0.0
$Usab.$	0.0	0.0	0.0	0.0
$Other$	0.0	$\frac{1 \cdot \frac{1}{2} \cdot (1+0)}{1} = 0.5$	$\frac{1 \cdot \frac{1}{2} \cdot (0+1)}{1} = 0.5$	0.0

3. Re-estimating weights

Table 8: Example of probability distributions ($p_{\mathbb{M}}(c|\boldsymbol{x})$) and predictions ($\arg\max_c p_{\mathbb{M}}(c|\boldsymbol{x})$) given by a model $\mathbb{M}$ for examples of Table 1.

		$p_{\mathbb{M}}(c\|\boldsymbol{x})$			
$\boldsymbol{x}$	Prediction	*Inst.*	*Req.*	*Usab.*	*Other*
$\boldsymbol{x}^1$	*Inst.*	0.7	0.1	0.0	0.2
$\boldsymbol{x}^2$	*Req.*	0.1	0.6	0.1	0.2
$\boldsymbol{x}^3$	*Usab.*	0.1	0.3	0.4	0.2
$\boldsymbol{x}^4$	*Req.*	0.2	0.5	0.3	0.0

Table 9: Computation of per-label weights (w_c^a, Eq. 7) for annotations of Table 1 and model results of Table 8.

Accuracy-based strategy:

c	L_1	L_2	L_3
Inst.	$\frac{1}{1} = 1.0$	0.0	$\frac{1}{1} = 1.0$
Req.	$\frac{1}{2} = 0.5$	$\frac{1}{1} = 1.0$	$\frac{2}{2} = 1.0$
Usab.	$\frac{0}{1} = 0.0$	$\frac{1}{2} = 0.5$	0.0
Other	0.0	$\frac{0}{1} = 0.0$	$\frac{0}{1} = 0.0$

Probability-based strategy:

c	L_1	L_2	L_3
Inst.	$\frac{0.7}{1} = 0.7$	0.0	$\frac{0.7}{1} = 0.7$
Req.	$\frac{0.6+0.3}{2} = 0.45$	$\frac{0.6}{1} = 0.6$	$\frac{0.6+0.5}{2} = 0.55$
Usab.	$\frac{0.3}{1} = 0.3$	$\frac{0.4+0.3}{2} = 0.35$	0.0
Other	0.0	$\frac{0.2}{1} = 0.2$	$\frac{0.2}{1} = 0.2$

Table 10: Computation with *accuracy-based strategy* of confusion-matrix weights ($W^a_{cc'}$, Eq. 8) for annotations of Table 1 and model results of Table 8.

L_1:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{1}{1} = 1.0$	0.0	0.0	0.0
$Req.$	0.0	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$	0.0
$Usab.$	0.0	$\frac{1}{1} = 1.0$	$\frac{0}{1} = 0.0$	0.0
$Other$	0.0	0.0	0.0	0.0

L_2:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	0.0	0.0	0.0	0.0
$Req.$	0.0	$\frac{1}{1} = 1.0$	0.0	0.0
$Usab.$	0.0	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$	0.0
$Other$	$\frac{1}{1} = 1.0$	0.0	0.0	$\frac{0}{1} = 0.0$

L_3:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{1}{1} = 1.0$	0.0	0.0	0.0
$Req.$	0.0	$\frac{2}{2} = 1.0$	0.0	0.0
$Usab.$	0.0	0.0	0.0	0.0
$Other$	0.0	0.0	$\frac{1}{1} = 1.0$	$\frac{0}{1} = 0.0$

Table 11: Computation with *probability-based strategy* of confusion-matrix weights ($W^a_{cc'}$, Eq. 8) for annotations of Table 1 and model results of Table 8.

L_1:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{0.7}{1} = 0.7$	$\frac{0.1}{1} = 0.1$	$\frac{0.0}{1} = 0.0$	$\frac{0.2}{1} = 0.2$
$Req.$	$\frac{0.1+0.1}{2} = 0.1$	$\frac{0.6+0.3}{2} = 0.45$	$\frac{0.1+0.4}{2} = 0.25$	$\frac{0.2+0.2}{2} = 0.2$
$Usab.$	$\frac{0.2}{1} = 0.2$	$\frac{0.5}{1} = 0.5$	$\frac{0.3}{1} = 0.3$	$\frac{0.0}{1} = 0.0$
$Other$	0.0	0.0	0.0	0.0

L_2:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	0.0	0.0	0.0	0.0
$Req.$	$\frac{0.1}{1} = 0.1$	$\frac{0.6}{1} = 0.6$	$\frac{0.1}{1} = 0.1$	$\frac{0.2}{1} = 0.2$
$Usab.$	$\frac{0.1+0.2}{2} = 0.15$	$\frac{0.3+0.5}{2} = 0.4$	$\frac{0.4+0.3}{2} = 0.35$	$\frac{0.2+0.0}{2} = 0.1$
$Other$	$\frac{0.7}{1} = 0.7$	$\frac{0.1}{1} = 0.1$	$\frac{0.0}{1} = 0.0$	$\frac{0.2}{1} = 0.2$

L_3:

$c \backslash c'$	$Inst.$	$Req.$	$Usab.$	$Other$
$Inst.$	$\frac{0.7}{1} = 0.7$	$\frac{0.1}{1} = 0.1$	$\frac{0.0}{1} = 0.0$	$\frac{0.2}{1} = 0.2$
$Req.$	$\frac{0.1+0.2}{2} = 0.15$	$\frac{0.6+0.5}{2} = 0.55$	$\frac{0.1+0.3}{2} = 0.2$	$\frac{0.2+0.0}{2} = 0.1$
$Usab.$	0.0	0.0	0.0	0.0
$Other$	$\frac{0.1}{1} = 0.1$	$\frac{0.3}{1} = 0.3$	$\frac{0.4}{1} = 0.4$	$\frac{0.2}{1} = 0.2$