An Investigation of Cross-Project Learning in Online Just-In-Time Software Defect Prediction – Supplementary Material

Sadia Tabassum
sxt901@cs.bham.ac.uk
University of Birmingham, UK

Leandro L. Minku∗
L.L.Minku@cs.bham.ac.uk
University of Birmingham, UK

Danyi Feng
danyi@ouchteam.com
Xiliu Tech, China

George Cabral
george.gcabral@ufrpe.br
Federal Rural University of Pernambuco, Brazil

Liyan Song
L.Song.1@cs.bham.ac.uk
University of Birmingham, UK

Algorithm 1 Ensemble approach

1: $S =$ stream of incoming changes from $n$ projects, $b =$ index identifying the test project, $w =$ waiting period

2: initialise ensemble model $M$ consisting of $n$ models $\{m_1, m_2, ..., m_n\}$

3: for each incoming change $x_p^t \in S$ do // $x_p^t$ is a change arriving from project $p$ at timestamp $t$

4: if $p = b$ then

5: $\hat{Y}$-List $\leftarrow$ get prediction $\hat{y}_i$ from each model in $M$

6: $\hat{y} =$ mean of stored prediction results in $\hat{Y}$-List

7: end if

8: store $x_p^t$ in a queue $WFL-Q_p$ for project $p$ // $WFL-Q_p$ is the queue of incoming changes of project $p$ waiting to be used for trained

9: for each model $m_p$ in $M$ do

10: for each change $q_i^t$ in $WFL-Q_p$ do

11: if a defect was linked to $q_i^t$ at a timestamp $\leq t$ then

12: create defect-inducing training_example for $q_i^t$

13: train($m_p$, training_example)

14: remove $q_i^t$ from $WFL-Q_p$

15: else

16: create a clean training_example for $q_i^t$

17: train($m_p$, training_example)

18: remove $q_i^t$ from $WFL-Q$

19: store training_example in $CL_p-H$ // $CL_p-H$ is a hash of clean training examples for project $p$

20: end if

21: end for

22: end for

23: if a defect was linked to a training_example in $CL_p-H$ before time $t$ then

24: Swap the label of training_example to defect-inducing

25: train($m_p$, training_example)

26: remove $h$ from $CL_p-H$

27: end if

28: end for

∗Corresponding author.
Figure 1: G-Mean for proprietary datasets through time using ORB. The vertical red bar indicates the last time step of the initial phase of the project.