Multi-Objective Software Effort Estimation

Federica Sarro*, Alessio Petrozziello† and Mark Harman*
University College London, London, United Kingdom*
University of Portsmouth, Portsmouth, UK†
f.sarro@ucl.ac.uk, alessio.petrozziello@port.ac.uk, mark.harman@ucl.ac.uk

ABSTRACT

This document is supplementary to the paper entitled "Multi-Objective Software Effort Estimation" [1], which is currently to appear in the proceedings of the 38th International Conference on Software Engineering (ICSE 2016).

1. RESULTS

1.1 RQ2. State of the Art Benchmark

In this Section we report the results of the Wilcoxon Test (based on MdAE) we performed in answer to RQ2 [1] (these results were omitted in our conference paper due to space limit).

We tested whether there was statistically significant difference between the Median of Absolute Error (MdAE) obtained with CoGEE and those obtained with CBR, LR, and CART. The results reported in Table 1 suggest that CoGEE is never worse than the state-of-the-art techqniues and it significanlty outperforms ($p \leq 0.001$) the them in 20 out 25 cases always with large effect size ($\hat{A}_{12} \geq 0.93$). Thus, confirming the results evaluated in terms of MAE and reported in our conference paper [1].

CoGEE vs	CBR1	CBR2	CBR3	LR	CART
China	<0.001 (1.00)	<0.001 (1.00)	< 0.001 (1.00)	1.00 (0.00)	<0.001 (1.00)
Desharnais	<0.001 (1.00)	<0.001 (0.97)	<0.001 (0.97)	0.009 (0.66)	<0.001 (0.97)
Finnish	<0.001 (1.00)	<0.001 (0.93)	<0.001 (1.00)	0.92 (0.40)	0.78 (0.60)
Maxwell	<0.001 (1.00)	<0.001 (1.00)	1.00 (0.10)	<0.001 (0.97)	1.00 (0.00)
Miyazaki	<0.001 (1.00)	<0.001 (1.00)	<0.001 (1.00)	<0.001 (1.00)	<0.001 (1.00)

Table 1: RQ2. Results of the Wilcoxon test (with $\hat{A}_{12}$ effect sizes in brackets) which compare the Median of the Absolute Errors (MdAE) for our algorithm, CoGEE, to those for the state-of-the-art techniques, CBR1–3, LR and CART.

1.2 RQ3. Multi-objective Benefit

In this section we report the figures of Pareto Fronts we produced in answer to RQ3 [1] (these results were omitted in our conference paper due to space limit).

Figure 1.2 shows the pareto fronts obtained in the 30 runs by using CoGEE and the single objective algorithms (i.e., GA-IC, GA-SAE).

We can observe that often none of the solutions produced by the single-objective algorithms lie on the pareto front, indicating that CoGEE allowed us to obtain the best results in terms of both CI and MAE on all the datasets we considered.

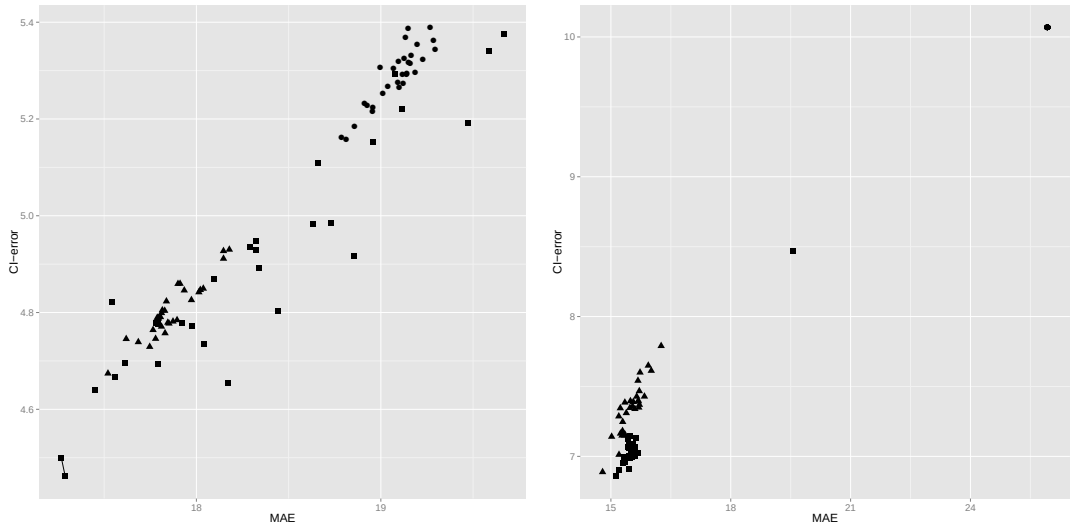
Figure 1.2 shows the pareto fronts obtained in the 30 runs by using CoGEE and the multi-objective algorithm NSGAII-UO, respectively. We can observe that the solutions provided by NSGAII-UO are far away from the optimal front for all the datasates considered. These results hold also when we use under and over estimates as evaluation criteria to compare the pareto fronts. (see Figure 1.2).

2. REFERENCES

[1] F. Sarro, A. Petrozziello, and M. Harman. Multi-objective effort estimation. In *Proceedings of ICSE2016, to appear*, 2016.

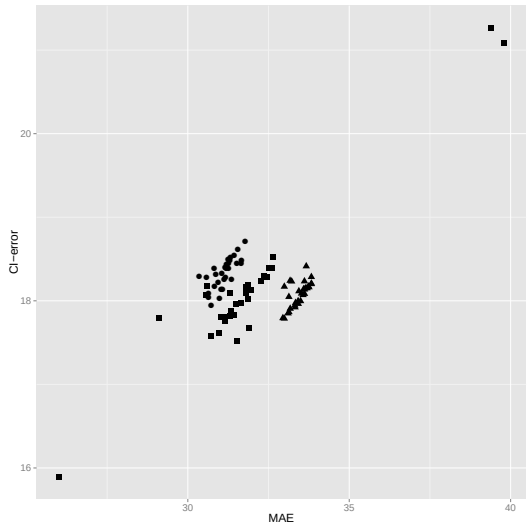
ICSE '16 May 14-22, 2016, Austin, TX, USA

ACM ISBN 978-1-4503-3900-1/16/05.
DOI: http://dx.doi.org/10.1145/2884781.2884830

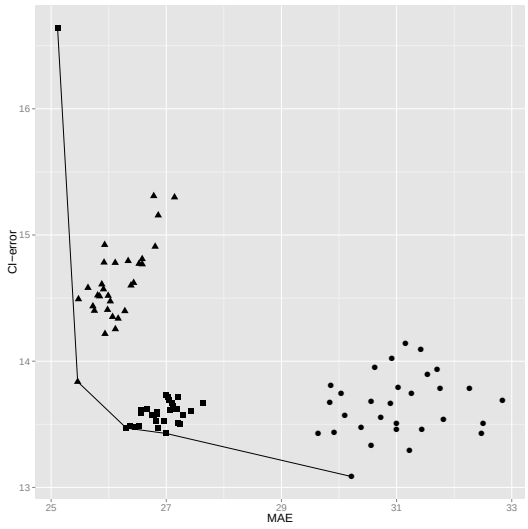


(a) China

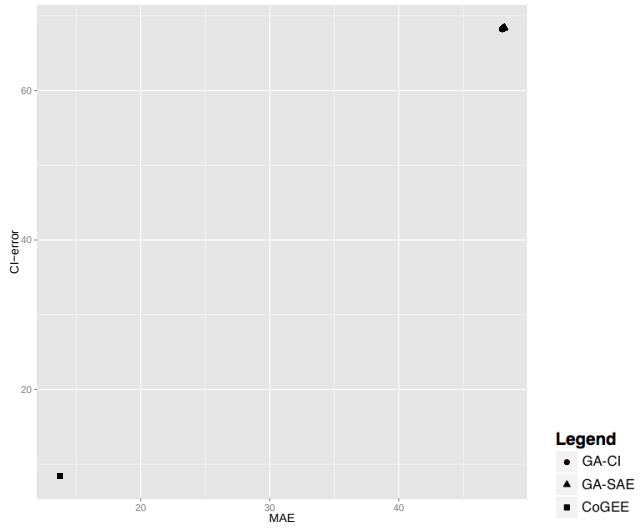
(b) Desharnais



(c) Finnish

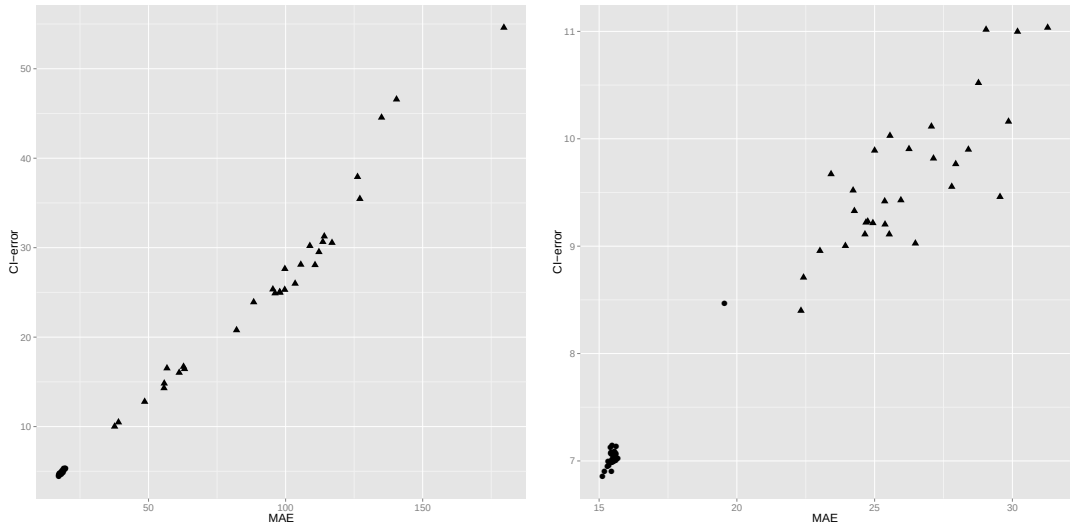


(d) Maxwell



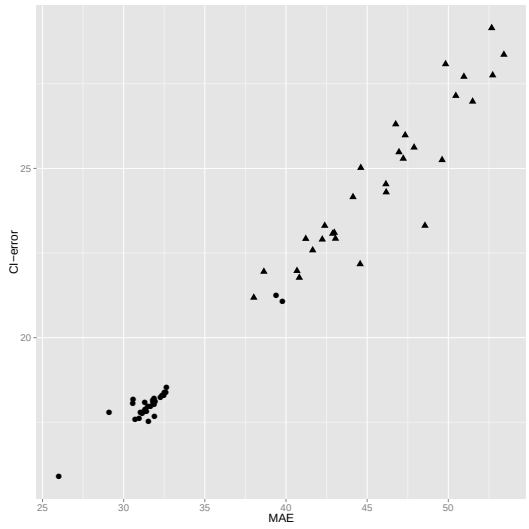
(e) Miyazaki

Figure 1: RQ3.1 Multi-objective Benefit. For each dataset we report the Pareto Fronts (in terms of CI and MAE) obtained by CoGEE, GA-SAE and GA-CI over 30 executions.

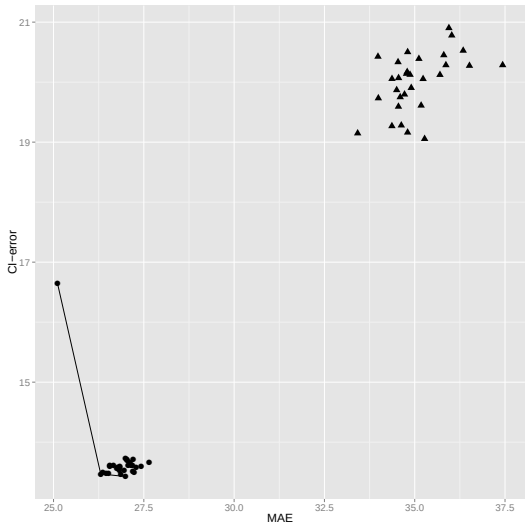


(a) China

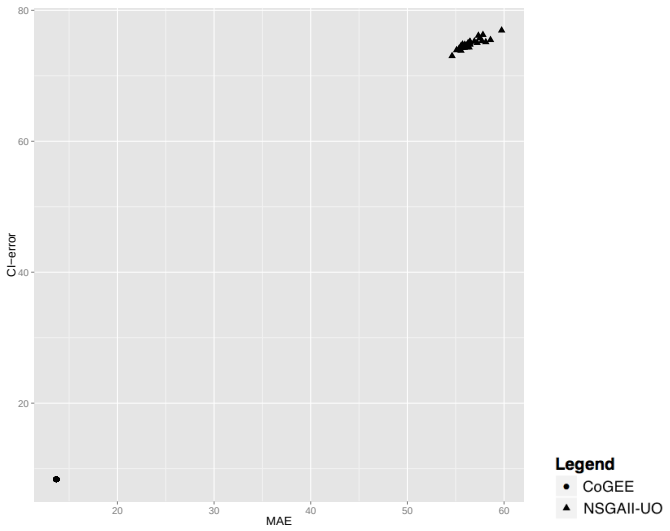
(b) Desharnais



(c) Finnish

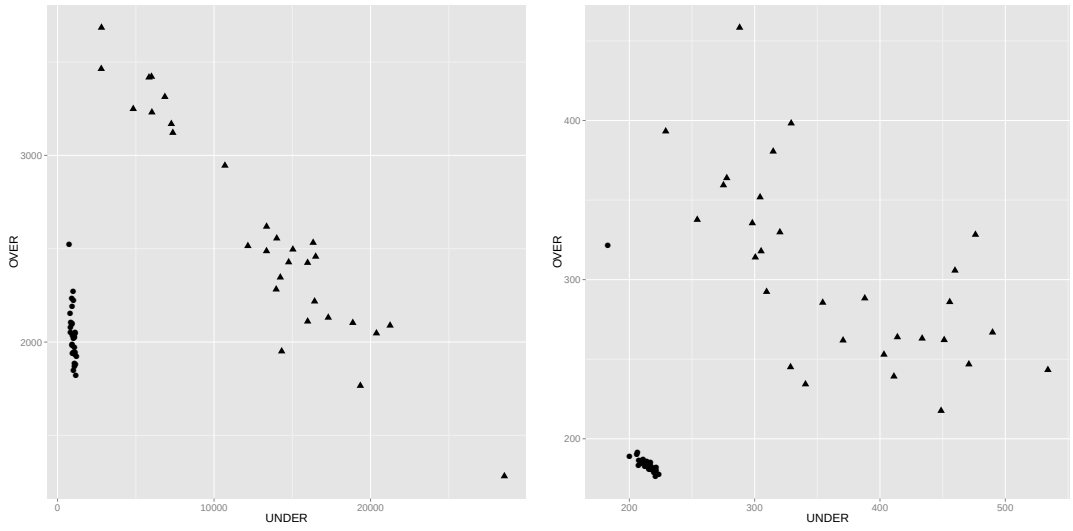


(d) Maxwell



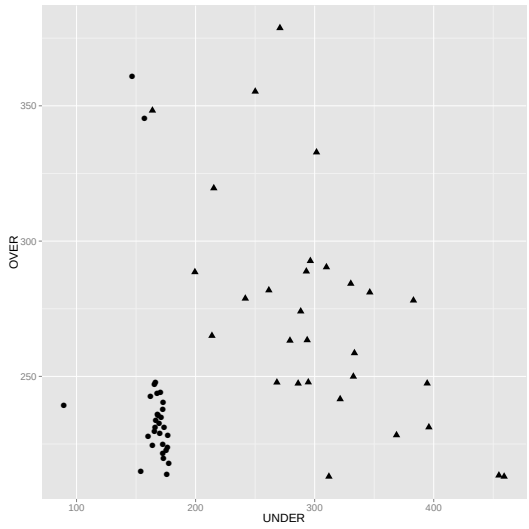
(e) Miyazaki

Figure 2: RQ3.2 Multi-objective Benefit. For each dataset we report the Pareto Fronts (in terms of CI and MAE estimates) obtained by CoGEE and NSGAII-UO over 30 executions.

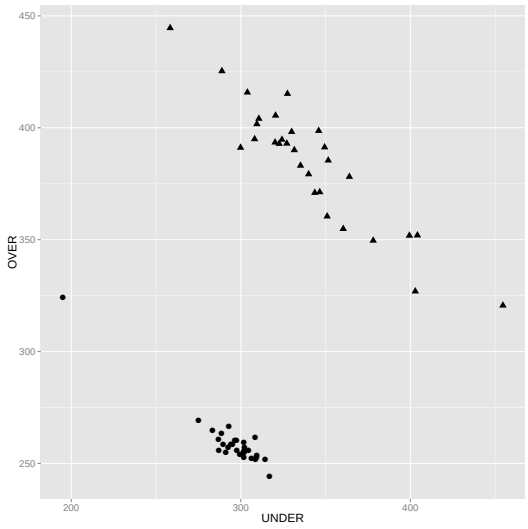


(a) China

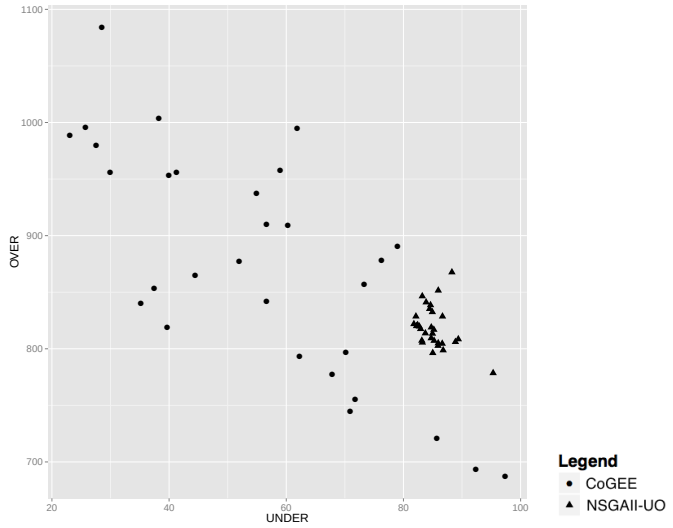
(b) Desharnais



(c) Finnish



(d) Maxwell



(e) Miyazaki

Figure 3: RQ3.2 Multi-objective Benefit. For each dataset we report the Pareto Fronts (in terms of over- and under- estimates) obtained by CoGEE and NSGAII-UO over 30 executions.