Energy Implications of Test Prioritisations in the Continuous Integration Process

What is the Energy Impact of Test Prioritization during CI Pipeline?

Rohan Cyr Lambert | R.C.Lambert@student.tudelft.nl | Faculty EEMCS, TU Delft

Supervisor: Xutong Liu | Responsible professor: Carolin Brandt

Keywords

- Test Selection (TS): Running only the tests that can be affected by the code changes.
- Test Prioritisation (TP): Ordering tests so those likely to fail are run earlier within the test suite. TP is almost always performed after TS.
- EnergiBridge (EB): A tool used to measure the energy consumption of a machine doing a task.

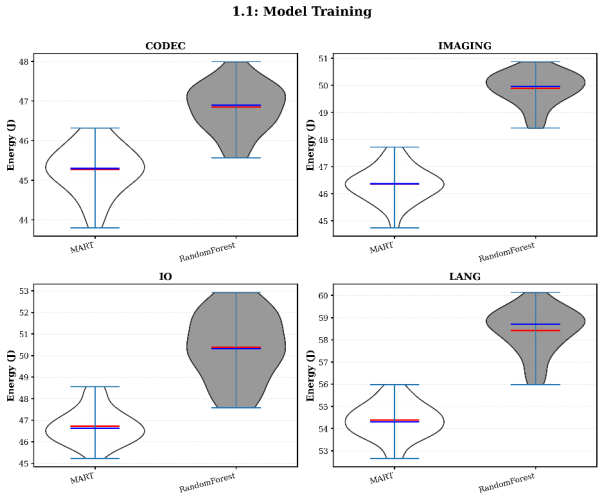
Background

- Running full pipeline is costly in Time and Energy especially when there are many commits[1].
- TP aims to reduce resource usage[2].
- Some research has been done on energy saved by Continuous Integration processes[3], but there is no data on specific parts like TP.
- This study reuses work from Bertolino et al (2020)[4]

Experiment

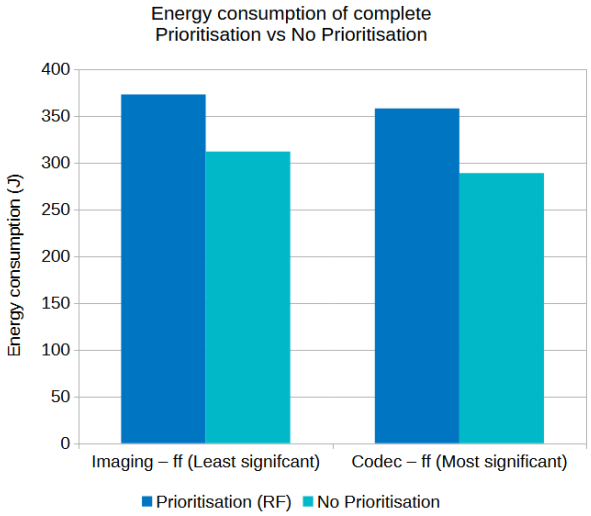
- Investigate TP in a contest similar to real world: TS and TP on 4 open-source Java projects.
- Train 2 Machine Learning models and rank the tests; run the tests ordered and non-ordered, both on commits with (nc) and without (ff) tests failing.
- Measure Energy usage with EB at every step and establish how much energy is saved/wasted.

Results



- Multiple Additive Regression Trees (MART) outperforms Random Forest both during training and Ranking.
-The Energy consumption for the 4 Java Commons projects was 45 to 60 Joules for training and ~30 Joules for ranking.

- Overall the Test Prioritisation rarely saved energy and when it did it was at most 20 Joules.
- The total energy consumption of the Prioritisation process was significantly higher than simply running the tests.



Discussion

- For the project on which this Prioritisation process has been ran, it has been much more costly in energy compared to what it saved. While MART outperformed RF in training and ranking resource usage, both prioritisation models failed to consistently make an impact on the energy usage of running the tests.
- Additionally, since in these type of projects, commits with failing tests are rare, the prioritisation is simply wasted energy with no benefits in the more than 97,5% of commits were there are no failing tests to be found.
- These results may be due to TS eliminating the need for TP on projects of this scale, this may be different for larger projects.

Conclusion

- Test Prioritisation is not worth investing resources into and will rather be a drain on Energy usage for projects of this scale utilising Test Selection.
- This Research recommends investigating into the energy usage of Test Selection and Test Prioritisation as stand-alone, as well as continue to similarly investigate the sub-processes of the CI pipeline, potentially on larger projects.

[1] A. Memon, Z. Gao, B. Nguyen, S. Dhanda, E. Nickell, R. Siemborski, and J. Micco, "Taming Google-scale continuous testing," in Proc. IEEE/ACM 39th Int. Conf. Software Engineering: Software Engineering in Practice Track (ICSE-SEIP), Buenos Aires, Argentina, May 2017, pp. 233–242, doi: 10.1109/ICSE-SEIP.2017.16.
[2] R. Pan, M. Bagherzadeh, T. A. Ghaleb, and L. Briand, "Test case selection and prioritization using machine learning: A systematic literature review," Empirical Software Engineering, vol. 27, no. 2, Art. no. 29, Dec. 2021, doi: 10.1007/s10664-021-10066-6.
[3] A. Zaidman, "An inconvenient truth in software engineering? The environmental impact of testing open source Java projects," in Proc. 5th ACM/IEEE Int. Conf. Automation of Software Test (AST), New York, NY, USA, Jun. 2024, pp. 214–218, doi: 10.1145/3644032.3644461.
[4] A. Bertolino, A. Guerriero, B. Miranda, R. Pietrantuono, and S. Russo, "Learning-to-rank vs ranking-to-learn: Strategies for regression testing in continuous integration," in Proc. ACM/IEEE 42nd Int. Conf. Software Engineering (ICSE), New York, NY, USA, Oct. 2020, pp. 1–12, doi: 10.1145/3377811.3380369.