Replicate the results

Hélia Galinier

March 27, 2025

1 Introduction

In this document, clear instructions as to how to replicate the results presented in the report will be provided.

2 Step by step process to replicate the results

Everything is going to be in the 'results' folder of the repository.

2.1 Extract the data for each dataset

This step is done with the file 'performance_comparison.py'.

To compute the results presented in the report, just run the code using the command : 'python performance_comparison.py' in the right folder or run the code with the IDE. The results for each dataset will be put into individual csv files with the name of the dataset followed by '_results.csv'. These files are stored in the 'performance_comparison' folder.

The results computed are the IDI ratio and runtime for both methods.

Some parameters can be changed. In particular, the number of runs done for each dataset is set to 10 and it is the number of runs that has been done to compute the results of the report. To change it, simply change the value of the variable n. The parameters of the tool and baseline functions can also be changed directly in them.

2.2 Extract the p-values and medians for the IDI ratio

This part has been done with the 'analyse_IDI_ratio.py' file. It just needs to be run with the usual command or with the IDE. The p-values, medians for each dataset and for the overall results are then printed on the console or terminal. It uses the data taken at the previous step. This file also provides figures for each dataset. These figures are stored in the folder 'IDI_ratio_plt' with the name of the figure in the following format : 'figure_' + name of the dataset + '.png'.

2.3 Extract the p-values and medians for the IDI per second

This part has been done exactly like the previous one. The file used is 'analyse_IDI_second.py'. The folder that stores the figures is 'IDI_second_plt'.

2.4 Extract the medians for the runtime

This part is done using the 'analyse_time.py' file. It just needs to be run with the usual command or IDE. The results are then printed on the console or terminal. These results are computed via the results extracted in the first step.