Configuration Report for the Solver RandomForest on the Training Instance Set Iris in *Sparkle*

Automatically generated by *Sparkle* (version: 0.9.3.2)

14th April 2025

1 Introduction

Sparkle [**?**] is a multi-agent problem-solving platform based on Programming by Optimisation (PbO) [**?**], and provides a number of effective algorithm optimisation techniques (such as automated algorithm configuration, portfolio-based algorithm selection, etc) to accelerate existing solvers.

This experimental report is automatically generated by *Sparkle*. This report presents experimental results on the scenario of configuring the solver RandomForest on the training instance set Iris.

2 Information about the Instance Set(s)

- Training set: **Iris**, consisting of 150 instances

3 Information about the Configuration Protocol

The configurator used by *Sparkle* in this scenario is SMAC2 (*Sequential Model-based Algorithm Configuration*) [**?**], and the version of SMAC2 used in *Sparkle* is 2.10.03.

During the configuration process, *Sparkle* performed 25 independent SMAC2 runs for configuring the solver RandomForest on the training instance set Iris; the configuration objective is accuracy:max (QUALITY); the whole configuration time budget is 600 seconds; the cutoff time for each run is 60 seconds.

Each independent run of SMAC2 attempts to find one optimised configuration. As a result, *Sparkle* would obtain 25 optimised configurations. Each of these was then evaluated on the entire training set, with one solver run per instance and a cutoff time of 60 seconds, and the configuration with the highest accuracy:max value was selected as the result of the configuration process.

4 Information about the Optimised Configuration

After the configuration process mentioned above, *Sparkle* obtained the optimised configuration. The details of the optimised configuration are described below.

1. bootstrap: True
2. criterion: gini
3. max_depth_type: None

4. max_features_special: sqrt
5. max_features_type: special
6. max_leaf_nodes_type: None
7. max_samples_type: None
8. min_impurity_decrease: 0.0
9. min_samples_leaf: 1
10. min_samples_split: 2
11. min_weight_fraction_leaf: 0.0
12. n_estimators: 100
13. oob_score: True
14. configuration_id: SMAC2_1744614958.9824817_10

5 Comparison between Configured Version and Default Version on the Training Instance Set

In order to investigate the performance on the training instance set, *Sparkle* ran the configured version of RandomForest and the default version of RandomForest on the training instance set. During this phase, each version performed one run per instance with a cutoff time of 60 seconds. The results are reported as follows.

- **RandomForest (configured)**, accuracy:max: 1.0
- **RandomForest (default)**, accuracy:max: 0.9666666666666668

The empirical comparison between the RandomForest (configured) and RandomForest (default) on the training set of Iris is presented in Figure 1.

Table 1 shows on how many instances the RandomForest (configured) and RandomForest (default) timed out (did not solve the instance within the cutoff time of 60 seconds) on the training set of Iris, as well as on how many instances both timed out.

configured	default	overlap
0	0	0

Table 1: Number of time-outs for RandomForest (configured), RandomForest (default), and for how many instances both timed out on the training set of Iris.

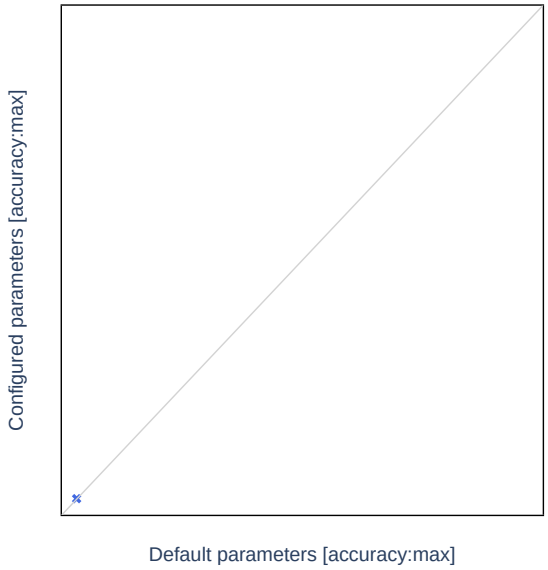

Figure 1: Empirical comparison between the RandomForest (configured) and RandomForest (default) on the training set of Iris.

References

[1] Holger H. Hoos. Programming by Optimization. *Communications of the ACM*, 55(2):70–80, 2012.

[2] Holger H. Hoos. Sparkle: A pbo-based multi-agent problem-solving platform. Technical report, Department of Computer Science, University of British Columbia, 2015.

[3] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In *Proceedings of the 5th International Conference on Learning and Intelligent Optimization (LION 5)*, pages 507–523, 2011.