Limitations of benchmark sets and landscape features for algorithm selection and performance prediction

Benjamin Lacroix
Robert Gordon University
Aberdeen, Scotland
b.m.e.lacroix@rgu.ac.uk

John McCall
Robert Gordon University
Aberdeen, Scotland
j.mccall@rgu.ac.uk

ABSTRACT

Benchmark sets and landscape features are used to test algorithms and to train models to perform algorithm selection or configuration. These approaches are based on the assumption that algorithms have similar performances on problems with similar feature sets. In this paper, we test different configurations of differential evolution (DE) against the BBOB set. We then use the landscape features of those problems and a case base reasoning approach for DE configuration selection. We show that, although this method obtains good results for BBOB problems, it fails to select the best configurations when facing a new set of optimisation problems with a distinct array of landscape features. This demonstrates the limitations of the BBOB set for algorithm selection. Moreover, by examination of the relationship between features and algorithm performance, we show that there is no correlation between the feature space and the performance space. We conclude by identifying some important open questions raised by this work.

CCS CONCEPTS

• **Computing methodologies → Heuristic function construction**; **Continuous space search**;

KEYWORDS

Continuous optimisation, numerical optimisation, benchmark functions, algorithm selection, exploratory landscape analysis

1 INTRODUCTION

Algorithm selection is based on the idea that landscape features can be used to efficiently select algorithms to solve a given problem. Recent research [4] [1] have achieved good results in that sense using standard benchmark such as BBOB. The limitation of such benchmarks have been demonstrated in multiple work, by their lack of diversity in the feature space [6] or in the performance space [2]. Hence, there is no evidence that the efficiency of using landscape features can be generalised to the whole problem space. In this paper, we propose to test the generalisation of the algorithm selection problem using landscape features. To do so, we develop a new set of benchmark problems using interpolated continuous optimisation problems (ICOP) [5] and assess the ability of the BBOB problem set to predict.

The framework for algorithm selection problem is based on the following elements [7]:

- $\mathcal{P}$ the problem space which represents a set of problems. In this paper, we consider two subsets problem spaces: (i) $P_{BBOB} \subset \mathcal{P}$ which contains five instances of the 24 BBOB problems in dimension 10 giving a set size of $|P_{BBOB}| = 120$. (ii) $P_{ICOP} \subset \mathcal{P}$, a set of 100 interpolated continuous problems (ICOP) [5] of dimension 10. The size of this set is $|P_{ICOP}| = 100$. The set of ICOP was generated using a genetic algorithm which evolves the a selection of seeds used to compose the ICOP. The objective is to mazimise the minimum distance between ICOPs in the feature space.

- $\mathcal{A}$ a set of algorithms designed to tackles problems from $\mathcal{P}$. In this paper we consider different configurations of differential evolution (DE)$A_{DE} \subset \mathcal{A}$ as the set of configurations $DE(F, Cr)$ with $F \in \{0.1, 0.2, ...1\}$ and $Cr \in \{0, 0.1, 0.2, ...1\}$. Giving us a total of 110 configurations ($|A_{DE}| = 110$).

- $\mathcal{F}$ the feature space, a set of features that can be evaluated for each elements in $\mathcal{P}$. Here we consider 20 features evaluated with the flacco package in R[3] using a sample size of 1000 solutions: **Fitness (cost) Distribution** (3), **Surrogate Models** (Meta-models) (9), **Information content features** (5), **Dispersion** : Dispersion Metric (DM) (1) and **Near Better Clustering** (2). The value of each feature is normalised between 0 and 1. The distance in $\mathcal{F}$ between a two problems in calculated by the Euclidean distance.

- $\mathcal{Y}$ the performance space which maps the performances of each algorithm in $\mathcal{A}$ In this paper, we define the performance space as the ranking of each algorithm with respect to each algorithms in $\mathcal{A}$. The distance in $\mathcal{Y}$ between two problems is calculated by the opposite of the spearman correlation of the performance of algorithms in A_{DE} [2].

2 ALGORITHM CONFIGURATION SELECTION

We evaluate the performances of each configurations in A_{DE} against each problem in P_{BBOB} and P_{ICOP} and rank them to obtain the mapping of both problem sets in the performance space. We then

use a case base reasoning approach based on the problems features to predict the best configuration for a given problem. We denote as $CBR_{P_{training}}(P_{testing})$ the performance of using CBR as a configuration selector for problems in $P_{testing}$ where the case base is composed of problems from $P_{training}$. First, the selection of DE configurations for the BBOB problems using other BBOB problems achieves a ranking of 22.36 giving a better mean ranking than any single configuration. This observation is in line with previous findings showing that features can be used to select the best algorithm/configuration in the algorithm space defined by the BBOB benchmark. However, when facing different problems such as the ones in P_{ICOP}, $CBR_{P_{BBOB}}$ obtains a mean ranking is of 60.98, which is above average. This indicates that $CBR_{P_{BBOB}}$ fails to generalise to very different problems in the feature space. The reciprocity here is valid, $CBR_{P_{ICOP}}$ obtains a mean ranking of 69.06 on BBOB problems. To prove that this lack of generalisation between the P_{ICOP} and P_{BBOB} is not due to the different nature of the problems, we use P_{ICOP} as the case base to select configurations for other problems in P_{ICOP}. Obtaining a mean ranking of 51.77. $CBR_{P_{ICOP}}$ also fails to select the best configurations even within the same problem class. With this analysis, we show that using landscape features to select algorithms to tackle a new problems should be applied with care. we see here that the efficiency of such strategy is highly subject to the knowledge base used and the new problems faced.

3 CORRELATION BETWEEN THE FEATURE SPACE AND THE PERFORMANCE SPACE

The idea of using the problem features to select the best algorithm to a given problem is based on the assumption that algorithms will have similar performance of similar problems. In other word, problems with similar footprint in the feature space will have similar footprint in the performance space. To verify this assumption, we estimate the correlation between the pair-wise distance in the feature space and in the performance space of each problems in P_{ICOP} and P_{BBOB}. In Figure 1, each point represents a pair of problems $\{p_i, p_j\}$ in P_{ICOP} and P_{BBOB}. The Spearman correlation for $p_i, p_j \in P_{BBOB}$ is 0.33 (blue points), for $p_i, p_j \in P_{ICOP}$ is 0.08 (red points), and for $p_i \in P_{BBOB}$ and $p_j \in P_{ICOP}$ is -0.09 (green points). Globally, we observe no correlation between distances in the feature space and distances in the performance space. This observation goes against the the assumption that landscape features can be used to select the best algorithm or predict an algorithm performance for a given problem. However, the fact that the correlation is higher within BBOB problems explains the possibility of using BBOB problems for that purpose. On the other hand, the lack of correlation within the ICOP set and between the ICOP and BBOB set explains the inability of selecting the best algorithm as shown in the previous section. These observations emphasize on the fact that landscape features cannot always be used for algorithm selection.

4 CONCLUSION

This work poses many questions on the methodology used for algorithm selection, algorithm performance prediction and consequently algorithm performance comparison. We showed here that the benchmark sets used offer a limited representation of the problem space which can limit the generalisation of future analysis on
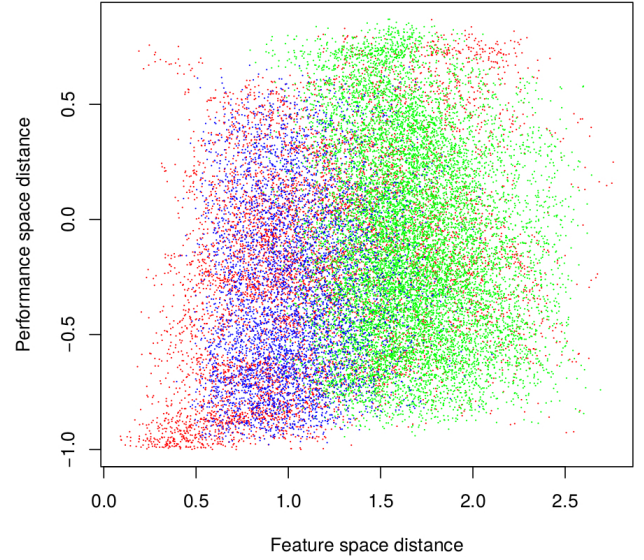
Figure 1: Distance in the feature space and performance space for problem pairs $\{p_i, p_j\}$ where $p_i, p_j \in P_{BBOB}$ • , $p_i, p_j \in P_{BBOB}$ •, $p_i \in P_{BBOB}$ and $p_j \in P_{ICOP}$ •
,

algorithm performance and algorithm selection. On the other hand, the results in this paper are bound by the set of landscape features used. It can thus be questioned if the lack of correlation between the feature space and the performance space is due to the choice in the landscape features used. Further analysis on the importance of each feature for that purpose is required, and possibly the addition of other features. Finally, the correlation analysis only between the feature space and the performance space is limited to linear relationship between features and performances. Advanced machine learning models could put in relief a more complex relationship.

REFERENCES

[1] Nacim Belkhir, Johann Dréo, Pierre Savéant, and Marc Schoenauer. 2016. Feature based algorithm configuration: A case study with differential evolution. In *International Conference on Parallel Problem Solving from Nature*. Springer, 156–166.

[2] Lee A Christie, Alexander EI Brownlee, and John R Woodward. 2018. Investigating benchmark correlations when comparing algorithms with parameter tuning. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 209–210.

[3] Pascal Kerschke. 2017. Comprehensive Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems Using the R-Package flacco. (2017).

[4] Pascal Kerschke and Heike Trautmann. 2017. Automated Algorithm Selection on Continuous Black-Box Problems By Combining Exploratory Landscape Analysis and Machine Learning. *arXiv preprint arXiv:1711.08921* (2017).

[5] Benjamin Lacroix, Lee A Christie, and John AW McCall. 2017. Interpolated continuous optimisation problems with tunable landscape features. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 169–170.

[6] Mario A. Muñoz and Kate A. Smith-Miles. 2016. Performance Analysis of Continuous Black-Box Optimization Algorithms via Footprints in Instance Space. *Evolutionary Computation* 0, 0 (2016), 1–26.

[7] Kate A. Smith-Miles. 2008. Cross-disciplinary perspectives on meta-learning for algorithm selection. *Comput. Surveys* 41, 1 (2008), 1–25. arXiv:1011.1669