Towards an algorithm selection standard: data format and tools

Lars Kotthoff[1]

The Algorithm Selection Problem is attracting increasing attention from researchers and practitioners from a variety of different backgrounds. After decades of fruitful applications in a number of domains, a lot of data has been generated and many approaches tried, but the community lacks a standard format or repository for this data. Furthermore, there are no standard implementation tools. This situation makes it hard to effectively share and compare different approaches and results on different data. It also unnecessarily increases the initial threshold for researchers new to this area.

In this talk, I will first give a brief introduction to the Algorithm Selection Problem and approaches to solving it [4, 3]. Then, I will present a standardized format for representing algorithm selection scenarios and a repository that contains a growing number of data sets from the literature, Aslib [1]. The format has been designed to be able to express a wide variety of different scenarios. In addition to encoding instance features and algorithm performances, there are facilities for providing feature costs, the status of algorithm execution and feature computations, cross-validation splits and meta-information. In addition to the data format itself, there is an R package that implements parsers and basic analysis tools. I will illustrate its usage through a series of examples.

I will further present LLAMA [2], a modular and extensible toolkit implemented as an R package that facilitates the exploration of a range of different portfolio techniques on any problem domain. It implements the algorithm selection approaches most commonly used in the literature and leverages the extensive library of machine learning algorithms and techniques in R. I will provide an overview of the architecture of LLAMA and the current implementation.

Leveraging the standard data format and the LLAMA toolkit, I will conclude this talk by presenting a set of example experiments that build and evaluate algorithm selection models. The models are created and evaluated with LLAMA on the problems in the algorithm selection benchmark repository. The results demonstrate the potential of algorithm selection to achieve significant performance improvements even through straightforward application of existing techniques.

Together, Aslib and LLAMA provide a low-threshold starting point for researchers wanting to apply algorithm selection to their domain or prototype new approaches. Both are under active development.

Joint work with Bernd Bischl, Pascal Kerschke, Marius Lindauer, Yuri Malitsky, Alexandre Fréchette, Holger Hoos, Frank Hutter, Kevin Leyton-Brown, Kevin Tierney, and Joaquin Vanschoren.

REFERENCES

[1] Bernd Bischl, Pascal Kerschke, Lars Kotthoff, Marius Lindauer, Yuri Malitsky, Alexandre Fréchette, Holger Hoos, Frank Hutter, Kevin Leyton-Brown, Kevin Tierney, and Joaquin Vanschoren, 'Aslib: A benchmark library for algorithm selection', *under review for AIJ*, (2014).

[2] Lars Kotthoff, 'LLAMA: leveraging learning to automatically manage algorithms', Technical Report arXiv:1306.1031, arXiv, (June 2013).

[3] Lars Kotthoff, 'Algorithm selection for combinatorial search problems: A survey', *AI Magazine*, (2014).

[4] Lars Kotthoff, Ian P. Gent, and Ian Miguel, 'An evaluation of machine learning in algorithm selection for search problems', *AI Communications*, **25**(3), 257–270, (2012).

[1] University College Cork, `larsko@4c.ucc.ie`