MULTI-FIDELITY EXPENSIVE BLACK-BOX MODELLING: AN EXAMINATION OF TRUST

NICOLAU ANDRÉS-THIÓ [ID]

In recent years, methods which tackle multi-fidelity expensive black-box (Mf-EBB) problems have received increasing attention due to their widespread presence in the industrial setting. Problems of this type are characterised by a lack of an analytical expression defining the relationship between design variables and chosen quality metrics. The key challenge is that knowledge of this relationship is limited to a small set of sample observations of varying quality. In the field of Mf-EBB, a problem instance consists of an expensive yet accurate source of information, and one or more cheap yet less accurate sources of information. The field aims to provide techniques either to accurately explain how decisions affect design outcome, or to find the best decisions to optimise design outcomes.

Surrogate modelling techniques, in particular, have successfully been applied to Mf-EBB problems in the past. These techniques are highly relevant when assessing the performance of a particular design which carries a high cost, as the overall cost can be mitigated via the construction of a model to be queried *in lieu* of the available high-cost source. The construction of these models can also simultaneously incorporate multiple information sources. Only in recent years, however, have researchers begun to explore the conditions under which these techniques are reliable. The existence of multiple sources, in particular, poses the question of which ones should be used when constructing a model.

The work of this thesis examines different aspects of Mf-EBB surrogate modelling methods. The main contributions centre around the concept of trust on an information source during model construction. Predicting whether incorporating certain datasets will increase or decrease model performance is no trivial matter, as will be shown

---

in this work. Novel test instances which are more varied than those found in the literature are proposed, as well as new features used to determine when the low-fidelity information source should be used. These are used to expose the criticality of a more heterogeneous test suite for algorithm assessment, as results are obtained which contradict existing literature guidelines. Next, the characterisation of harmful low-fidelity sources is approached as an algorithm selection problem. By performing this analysis using only the limited data available to train a surrogate model, new guidelines are provided that can be directly used in an applied industrial setting. Finally, the characterisation is used to create novel techniques for model creation with budget constraints which adaptively select the model to be trained based on the available data. These techniques are shown to be safer than traditional approaches, and highlight the benefits that result from the insights found in this thesis.

Some of this research has been published in [1–3].

Acknowledgements

References

[1] N. Andrés-Thió, M. A. Muñoz and K. Smith-Miles, 'Bi-fidelity surrogate modelling: showcasing the need for new test instances', *INFORMS J. Comput.* **34**(6) (2022), 3007–3022.
[2] N. Andrés-Thió, M. A. Muñoz and K. Smith-Miles, 'Characterising harmful data sources when constructing multi-fidelity surrogate models', *Artificial Intelligence* **336** (2024), Article no. 104247.
[3] N. Andrés-Thió, M. A. Muñoz and K. Smith-Miles, 'Methodology and challenges of surrogate modelling methods for multi-fidelity expensive black-box problems', *ANZIAM J.* **66**(1) (2024), 35–61.

NICOLAU ANDRÉS-THIÓ, School of Mathematics and Statistics,
University of Melbourne, Parkville, Victoria 3010, Australia
and
ARC Training Centre in Optimisation Technologies, Integrated Methodologies, and Applications (OPTIMA),
Parkville, Victoria 3010, Australia
e-mail: nandresthio@gmail.com