

INTERNSHIP PROPOSAL

Improvement of the distance to deal with categorical variables in a surrogate-based evolutionary algorithm

Cenaero, located in Gosselies (Belgium), is a private non-profit applied research center providing companies involved in a technology innovation process with numerical simulation methods and tools to invent and design more competitive products. Internationally recognized, in particular through its research partnership with Safran, Cenaero is mainly active in the aerospace (with an emphasis on turbomachinery), process engineering, energy and building sectors.

Cenaero provides expertise and engineering services in multidisciplinary simulation, design and optimization in the fields of mechanics (fluid, structure, thermal and acoustics), manufacturing of metallic and composite structures as well as in analysis of in-service behavior of complex systems and life prediction. Cenaero also provides software through its massively parallel multi-physics platform Argo, its manufacturing process simulation and crack propagation platform Morfeo and its design space exploration and optimization platform Minamo.

Cenaero operates the Tier-1 Walloon supercomputing infrastructure Lucia with 4 Pflops peak performance and was ranked 245th on the November 2022 Top500 List (see tier1.cenaero.be for details).

Within Cenaero, the Machine Learning and Optimization group is dedicated to the development of algorithms and methods to address complex industrial design cases, with several achievements in aeronautics in particular [1-2]. It incorporates the Minamo team, dedicated to the development of Cenaero's in-house multi-disciplinary optimization platform. Although computing power has increased dramatically in the last decades, computational burden is still an issue as more and more complex simulation analyses are required in industrial design processes. Aiming to tackle this numerical challenge, Minamo provides efficient online Surrogate-Based Optimization (SBO) methods, based on evolutionary algorithms, allowing to quickly gain insight into the design space, to quantitatively identify key factors and trends and to automatically find innovative design options. It implements several variants of mono- and multi-objective evolutionary strategies, efficiently coupled in an online framework (i.e with continuous enrichment of the construction support along the design iterates) to surrogate models.

Context

A classical SBO approach consists of several major components, as shown in Figure 1.

The first step relates to the Design of Experiments (DoE) which provides data for the initial surrogate model training. The second step consists in the training of the surrogate model(s). The third aspect of such optimization techniques is the choice of the model updating strategy. The last step is the evaluation of the stopping criterion to determine whether a new iterate is needed or not. In this context, a major challenge is to efficiently handle the presence of mixed variables, i.e. variables of different natures (continuous, integer, discrete and categorical), that can be found in numerous real-world problems, such as :

- number of components (integer/discrete variables);
- different materials (categorical variables with no intrinsic order);
- geometrical variables (continuous variables in general).

Since the construction of the surrogate models and the implemented Evolutionary Algorithm (EA) require the comparison of individuals with variables of different types, the best approach is to define a specific distance, as mentioned in [3]. One way to handle applications with both continuous and categorical variables is by means of a heterogeneous distance function that uses specific implementations for the different types of variables. One approach that has been used in Minamo is based on the overlap metric for categorical variables and the normalized Euclidean distance for continuous variables (real, integer, and discrete ones).

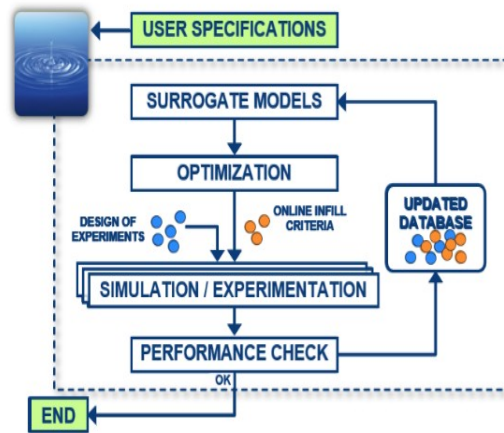


Figure 1: Online SBO scheme

Objective

The aim of this work is to:

- Investigate and implement other promising counterparts to the overlap metric. For instance : Inverse Occurrence Frequency (IOF), Occurrence Frequency (OF) and Burnaby metrics, as described in [4];
- Compare the existing distance with its selected competitors in a surrogate-based optimization strategy on different mathematical problems.

Depending on the progress and results obtained during this internship, these strategies may be tested on a realistic benchmark related to wing design.

The candidate will join the Machine Learning & Optimization group.

Profile

- Required: Master's student in Mathematics, Engineering or Computer Science.
- Languages: English and/or French.
- Prerequisites: Good programming skills as well as good mathematical background. Working knowledge of Linux and Python are valuable assets.
- Motivation, creativity and team spirit!

Duration

The length of the internship can vary from 3 months to 6 months, depending on your university or school regulations.

Contact

If you are interested in the topic, please send a resume and a cover letter quoting the reference number of the offer to rh_be-ip-2025-003@cenaero.be.

References

- [1] Baert, L., Chérière, E., Sainvitu, C., Lepot, I., Nouvellon, A., Leonardon, V. *Aerodynamic Optimisation of the Low Pressure Turbine Module: Exploiting Surrogate Models in a High-Dimensional Design Space*. Journal of Turbomachinery. 142:1-24 (2020).
- [2] Beaucaire, P., Beauthier, C., Sainvitu, C. *Multi-point infill sampling strategies exploiting multiple surrogate models*. GECCO '19: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1559-1567 (2019).
- [3] Wilson, D.R., Martinez, T.R.: *Improved Heterogeneous Distance Functions*. Journal of Artificial Intelligence Research 6, 1–34 (1997)
- [4] Boriah, S., Chandola, V., and Kumar, V., *Similarity Measures for Categorical Data: A Comparative Evaluation*. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pages 243–254. Society for Industrial and Applied Mathematics, (2008).