Initial Algorithm Selection, infrastructure and hardware analysis

HORIZON-CL4-2023-TWIN-TRANSITION-01-02.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union

Initial Algorithm Selection using artificial data and Infrastructure Analysis

Inputs, Outputs and Expected Outcomes for the algorithm
Initial Infrastructure Considerations

In the **FLASH-IANUS platform**, the machine learning algorithm will be key in optimising processing regimes by predicting the best-fitted combination of parameters to produce high-quality outcomes. The selected algorithm will be trained and will rely on a range of material property data, machine laser type, and the environmental footprint of the laser processes to replace time-consuming trial-and-error methods with data-driven insights to improve production efficiency, reduce operational costs and environmental impact.

As part of these goals, **Cosmos Thrace's (COS) team** has been tasked with the algorithm selection and consequent implementation to power the predictive capabilities of the platform. Following the earlier work on defining the relevant input types and features, the expected outputs, and performance goals, the team has now moved and made significant progress on evaluating and selecting the initial modelling implementation approaches and the type of algorithm. The models' success will ultimately be measured against several key outcomes, once real data becomes available: improved production efficiency (e.g., increased throughput, optimised resource usage), increased cost efficiency (e.g., minimised costs related to setup, operation, materials, maintenance), improved sustainability (e.g., reduced environmental impact through efficient energy and material usage).

Algorithm implementation models and selection

COS has focused on two main modelling strategies to determine how to best train and implement the algorithm: **Forward Modelling (with optimisation)** and **Inverse Modelling**. All initial screenings and tests were run with dummy data, which is removed from the models' real-world applications but can still be utilised to show apparent patterns and aid in constructing the characteristics of each algorithm option, preparing the test versions for historical data trials.

In the forward approach, the model is trained to predict quality criteria based on processing parameters. The method allows for a transparent optimisation process and offers significant flexibility and control. During training, initial input values and constraints are provided to guide the optimisation process. Within this approach, two implementation strategies were explored. The first involves having **separate models for each quality criterion**. This allows us to use the inherent flexibility of the approach as different algorithms and hyperparameters can be chosen to suit the specific characteristics of each output criterion, but may limit the model's ability to capture potential correlations between quality criteria. The second strategy uses a single multi-output model - the difference is that this one model is used for all hyperparameters, which reduces the overall use case prediction flexibility. Nevertheless, this unification would allow the model to learn shared representations and dependencies among the targets.

Inverse modelling, in contrast, trains the model to map the desired quality criteria directly to the corresponding processing parameters. This allows for faster predictions, as the optimisation loop is skipped. As such, it will rely on parameter constraints as it will not be as grounded in the relevant use cases.

Specific models were tested for both approaches: random forests, gradient boosted forests, multilayer perceptron (MLP), and custom neural networks (NNs).

With historical data being available, the team has started benchmarking the models. The next steps will be to evaluate the performance of each model, assessing accuracy and robustness, as well as computational efficiency, to finalise the selection. The selection needs to be further verified once real-world data becomes available, as the FLASH system will bring its processing specifications.

Infrastructure and hardware analysis

In parallel with modelling work, COS is also conducting a focused infrastructure and hardware assessment to understand the computational load the shortlisted Machine learning models will place on the FLASH-IANUS platform. The investigation delves into CPU, GPU and memory-intensive stages of training and real-time inference and checks the feasibility of future scaling strategies. This ensures the underlying hardware stack can sustain high-frequency predictions without bottlenecks while leaving headroom for future algorithm upgrades.