MASIF: Meta-learned Algorithm Selection using Implicit Fidelity Information

Tim Ruhkopf[1] **Aditya Mohan**[1] **Difan Deng**[1] **Alexander Tornede**[1] **Frank Hutter**[2]
Marius Lindauer[1]

[1]Institute of Artificial Intelligence, Leibniz University Hannover
[2]Machine Learning Lab, Albert-Ludwigs University Freiburg, Bosch Center for Artificial Intelligence

Abstract Selecting a well-performing algorithm for a given task or dataset can be time-consuming and tedious, but is crucial for the successful day-to-day business of developing new AI & ML applications. Algorithm Selection (AS) mitigates this through a meta-model leveraging meta-information about previous tasks. However, most of the available AS methods are error-prone because they characterize a task by either cheap-to-compute properties of the dataset or evaluations of cheap proxy algorithms, called landmarks. In this work, we extend the classical AS data setup to include multi-fidelity information and empirically demonstrate how meta-learning on algorithms' learning behaviour allows us to exploit cheap test-time evidence effectively and combat myopia significantly. We further postulate a budget-regret trade-off w.r.t. the selection process. Our new selector MASIF is able to jointly interpret online evidence on a task in form of varying-length learning curves without any parametric assumption by leveraging a transformer-based encoder. This opens up new possibilities for guided rapid prototyping in data science on cheaply observed partial learning curves.

1 Broader Impact Statement

After careful reflection, the authors have determined that this work presents no notable negative impacts on society or the environment.

2 Submission Checklist

1. For all authors…

   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Yes. The main claims can be found at the end of the introductory section 1 under "contributions" and reflect the essence of our work.

   (b) Did you describe the limitations of your work? [Yes] Yes, but we formulate them as future work in section 7.

   (c) Did you discuss any potential negative societal impacts of your work? [No] As the above broader impact statement conveys, we have no such concerns.

   (d) Have you read the ethics author's and review guidelines and ensured that your paper conforms to them? `https://automl.cc/ethics-accessibility/` [Yes] We have read the guidelines and conform with them.

2. If you are including theoretical results…

   (a) Did you state the full set of assumptions of all theoretical results? [N/A] We provide empirical work.

(b) Did you include complete proofs of all theoretical results? [N/A] We provide empirical work.

3. If you ran experiments…

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results, including all requirements (e.g., `requirements.txt` with explicit version), an instructive `README` with installation, and execution commands (either in the supplemental material or as a URL)? [Yes] We provide a GitHub repository that can be found at https://github.com/automl/masif, that includes all of the code and necessary instructions to install the environment and to reproduce our experiments.

(b) Did you include the raw results of running the given instructions on the given code and data? [No] The raw results data is not included.

(c) Did you include scripts and commands that can be used to generate the figures and tables in your paper based on the raw results of the code, data, and instructions given? [Yes] The code for generating the plots given the raw data is included in masif/post_processing.

(d) Did you ensure sufficient code quality such that your code can be safely executed and the code is properly documented? [Yes] Due to rapid development, the code can probably improve in the documentation.

(e) Did you specify all the training details (e.g., data splits, pre-processing, search spaces, fixed hyperparameter settings, and how they were chosen)? [Yes] We provide a bash script, detailing all necessary commands to reproduce our experiments. They are included in the bashfiles folder. Each script contains the necessary commands to execute the hydra pipeline with the necessary variations to the base config file found in configs/base. The configs folder includes all possible variations of hyperparameter settings, including data pre-processing and splits. The choice of hyperparameters is motivated in the ablations in Appendix A.4

(f) Did you ensure that you compared different methods (including your own) exactly on the same benchmarks, including the same datasets, search space, code for training and hyperparameters for that code? [Yes] Indeed, we have a detailed experiment section 4.1-4.3, detailing the protocol, the baselines and benchmarks used explicitly and which of the baselines can be applied to which benchmark given the requirements of both.

(g) Did you run ablation studies to assess the impact of different components of your approach? [Yes] The details can be found in Appendix A4.

(h) Did you use the same evaluation protocol for the methods being compared? [Yes] All of the methods are evaluated on the same Slice Evaluation Protocol we introduce in section 4.1.

(i) Did you compare performance over time? [N/A]

(j) Did you perform multiple runs of your experiments and report random seeds? [Yes] We used five seeds and ten folds for each benchmark. A crucial detail regarding the split in the algorithms' learning curves is described in the dataset section 4.3.

(k) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] We report confidence bands over the seeds.

(l) Did you use tabular or surrogate benchmarks for in-depth evaluations? [Yes] We relied on tabular benchmarks such as Task Set and LCBench. The details of which can be found in Appendix A.1

(m) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] All the experiments are executed on 4 Intel Xeon E5 cores with 8000MB RAM, we did however not keep track of the invested total amount of compute.

(n) Did you report how you tuned hyperparameters, and what time and resources this required (if they were not automatically tuned by your AutoML method, e.g. in a NAS approach; and also hyperparameters of your own method)? [Yes] We added ablations, that support our manual search. We refrained from more elaborated approaches, as data was already scarce due to the expense of acquiring the data for our meta-learner.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

   (a) If your work uses existing assets, did you cite the creators? [Yes] All baselines and benchmarks are cited sufficenctly.

   (b) Did you mention the license of the assets? [No] Since they are open-source.

   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We added the Scikit-CC18 dataset as supplementary. The synthetic dataset can be recreated from our code.

   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] Our experiments were conducted on publicly available datasets and we did not introduce new datasets.

   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] We do not make use of personal data.

5. If you used crowdsourcing or conducted research with human subjects...

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] We do not make use of human subjects.

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A] We do not make use of human subjects.

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A] We do not make use of human subjects.