Towards explainable meta-learning

Katarzyna Woźnica

Meta-learning, or *learning to learn*, is the science of systematically observing how different machine learning approaches perform on a wide range of learning tasks, and then learning from this experience, or *meta-data*, to learn new tasks much faster than otherwise possible.

Vanschoren, 2018

J. Schmidhuber, 1987

Overviews

- Timothy Hospedales, Antreas Antoniou, Paul Micaelli, Amos Storkey
  Meta-Learning in Neural Networks: A Survey (2020)
- Vanschoren, Joaquin *Meta-Learning: A Survey (2018)*

Meta-Learning: Minimise loss over a task distribution wrt meta-representation ω.

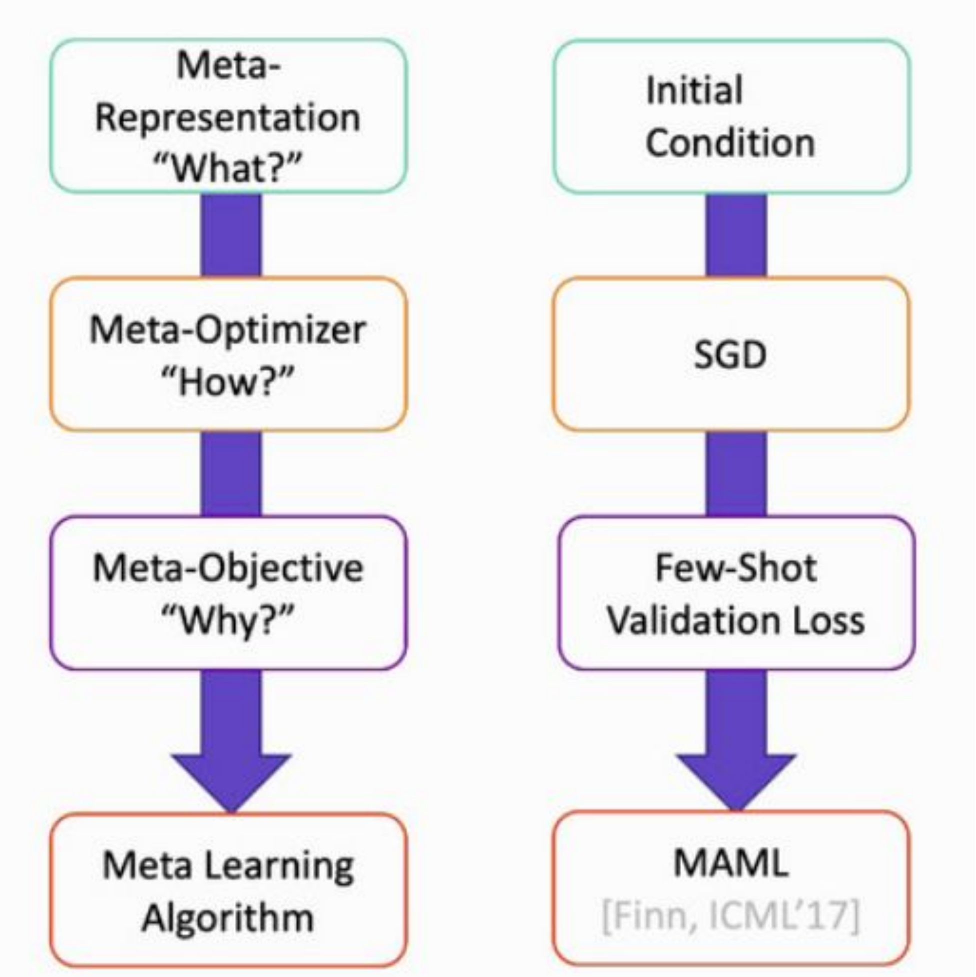
$$\min_{\omega} \mathbb{E}_{D \sim p(D)} L(D; \omega)$$

Meta-Training, Bi-level optimization view:
Outer: Train the algorithm ω
Inner: Train the model θ conditional on algorithm

$$\omega^* = \arg\min_{\omega} \sum_{t} L\left(D_t^{val}; \theta_t^*, \omega\right)$$

$$\text{s.t. } \theta_t^* = \arg\min_{\theta} L(D_{t,}^{trn}; \theta_t, \omega)$$

Meta-Testing: Deploy on a new task
$$\theta^* = \arg\min_{\theta} L(D_{novel}; \theta, \omega)$$



Hospedales T.,Presentation on Meta-Learning Workshop NiPS 2020

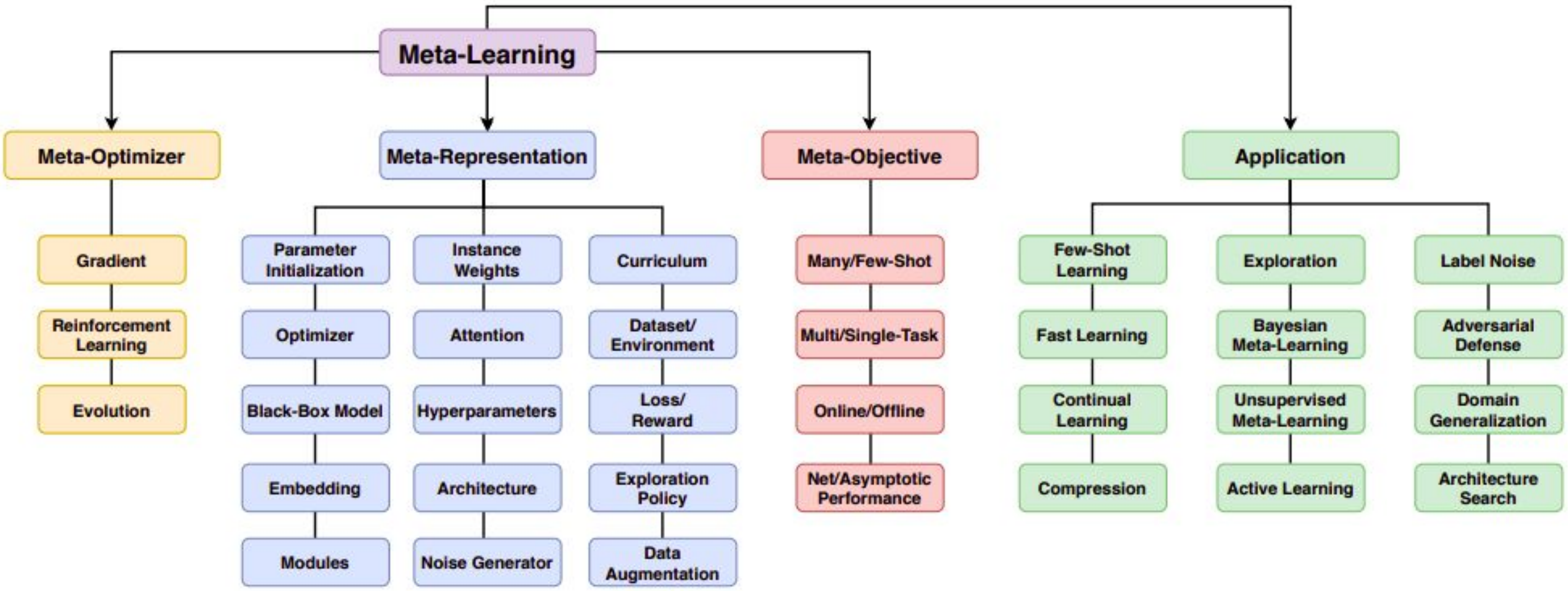


Fig. 1. Overview of the meta-learning landscape including algorithm design (meta-optimizer, meta-representation, meta-objective), and applications.

Hospedales, 2020

Meta-Representation	Meta-Optimizer		
	Gradient	RL	Evolution
Initial Condition	[16], [79], [88], [102], [166], [166]–[168]	[169]–[171] [16], [63], [64]	[172], [173]
Optimizer	[19], [94] [21], [39], [79], [106], [107], [174]	[81], [93]	
Hyperparam	[17], [69] [71]	[175], [176]	[173] [177]
Feed-Forward model	[38], [45], [86], [110], [178], [179] [180]–[182]	[22], [114], [116]	
Metric	[20], [90], [91]		
Loss/Reward	[42], [95] [127] [124]	[126] [121], [183] [124]	[123] [23] [177]
Architecture	[18] [135]	[26]	[25]
Exploration Policy		[24], [184]–[188]	
Dataset/Environment	[156] [159]	[162]	[163]
Instance Weights	[151], [152], [155]		
Feature/Metric	[20], [90]–[92]		
Data Augmentation/Noise	[145] [119] [189]	[144]	[146]
Modules	[140], [141]		
Annotation Policy	[190], [191]	[192]	

TABLE 1

Research papers according to our taxonomy. We use color to indicate salient meta-objective or application goal. We focus on the main goal of each paper for simplicity. The color code is: sample efficiency (red), learning speed (green), asymptotic performance (purple), cross-domain (blue).

Hospedales, 2020

CHALLENGES AND OPEN QUESTIONS

- Diverse and multi-modal task distributions: The difficulty of fitting a meta-learner to a distribution of tasks p(T) can depend on its width.
- Meta-generalization:
  - generalizing from meta-train to novel meta-test tasks drawn from p(T).
  - generalizing to meta-test tasks drawn from a different distribution than the training tasks.
  - which kinds of meta-representations tend to generalize better under certain types of domain shifts.
- Task families
- Computation Cost & Many-shot

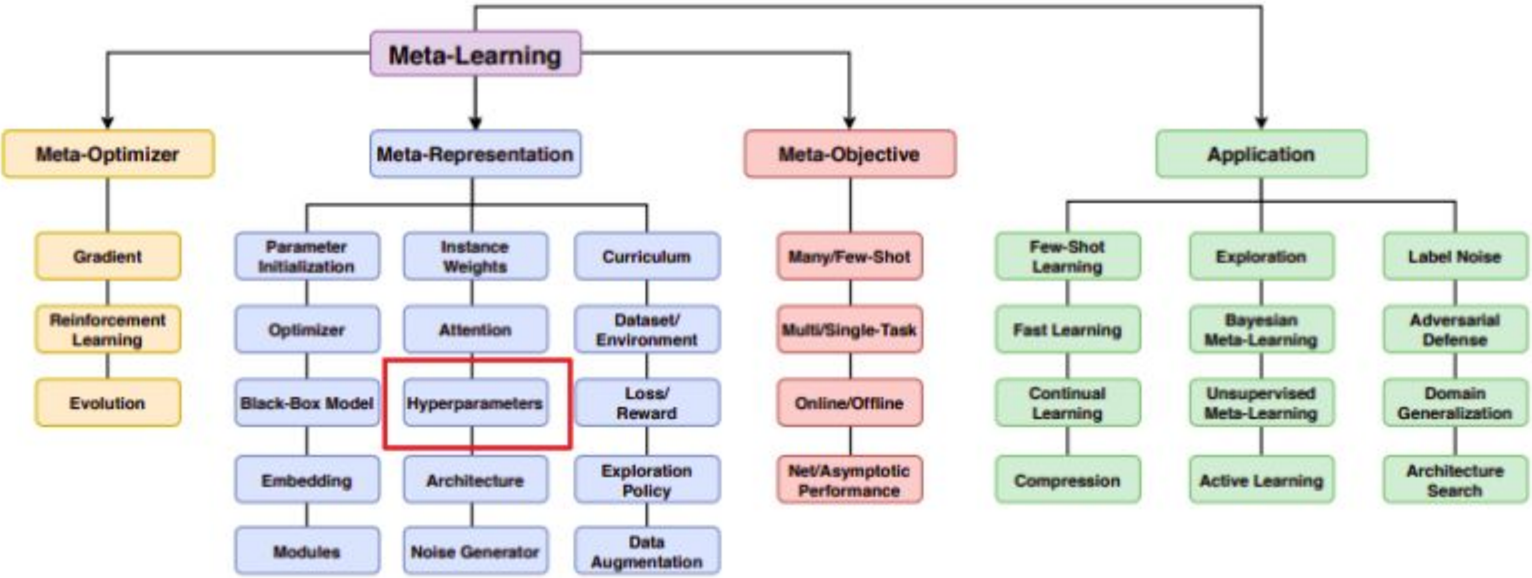
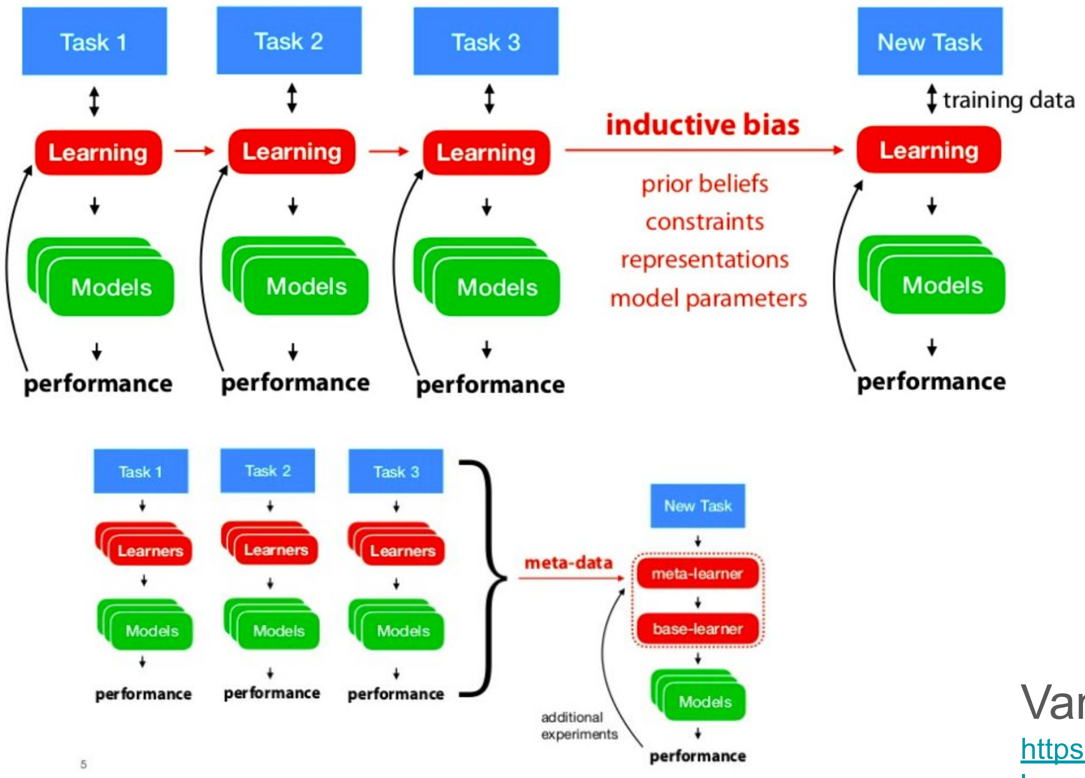


Fig. 1. Overview of the meta-learning landscape including algorithm design (meta-optimizer, meta-representation, meta-objective), and applications.

Hospedales, 2020

Vanschoren categorization:



- Learning from Model Evaluations
- Learning from Task Properties

Vanschoren, NIPS 2018
https://www.slideshare.net/JoaquinVanschoren/learning-how-to-learn-127759100

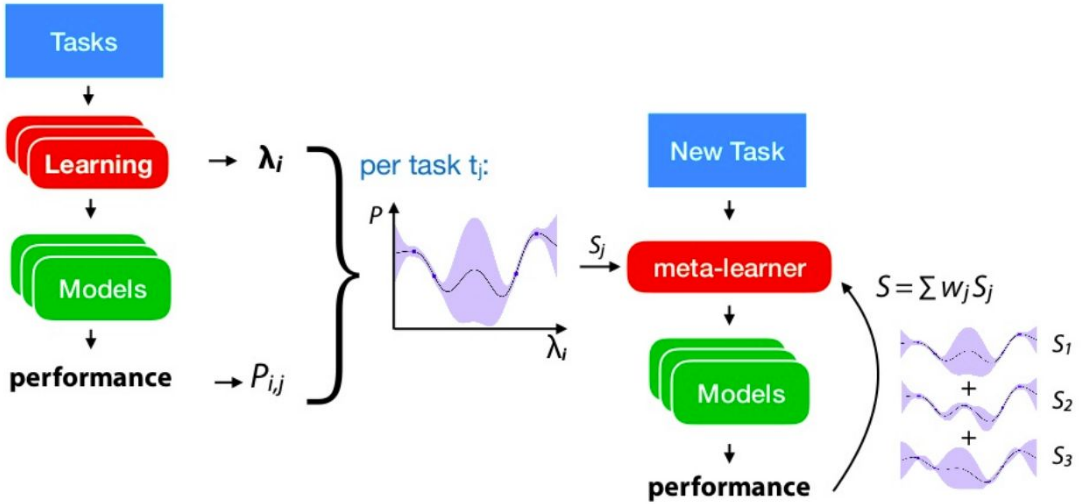
https://www.youtube.com/watch?v=0eBR8a4MQ30

Meta-features

- Hand crafted: simple, statistical, information-theoretic (e.g. entropy, mutual information), model-based, landmarkers,

- learning meta-features: with Siemense network (J. Kim, *Learning to warm-start Bayesian hyperparameter optimization. 2017)*

Surrogate model transfer:



Vanschoren, NIPS 2018

- M. Wistuba, N. Schilling, and L.Schmidt-Thieme. *Scalable Gaussian process-based transfer surrogates for hyperparameter optimization* (Machine Learning 2018)
- M. Feurer, B. Letham, and E. Bakshy. *Scalable meta-learning for Bayesian optimization.* (2018)
- G.Manolache J. Vanschoren. *Meta-Learning for Algorithm and Hyperparameter Optimization with Surrogate Model Ensembles* (2019)

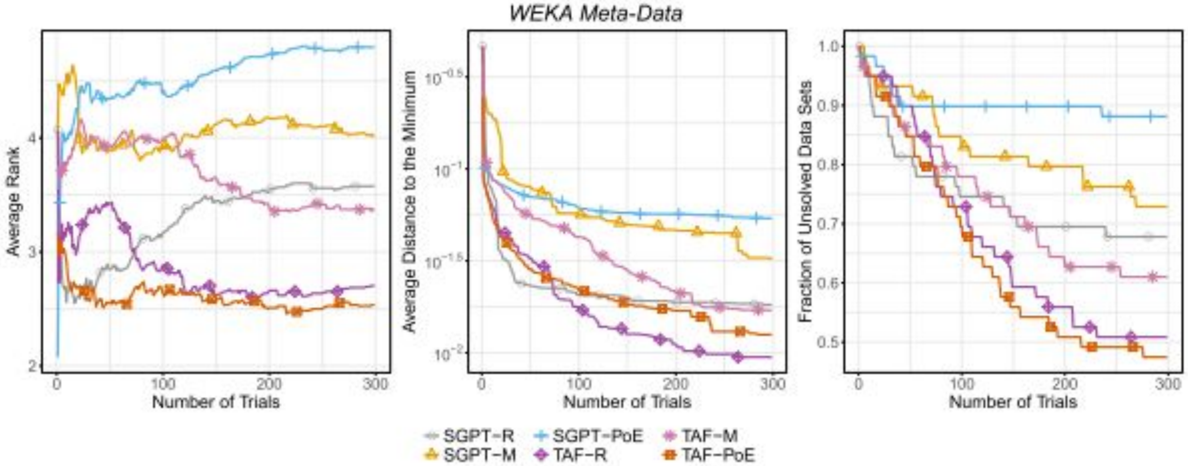


Fig. 10 TAF provides a clear improvement over SGPT thanks to its adaptive use of meta-data and better way of dealing with different data set scales

Scalable meta-learning for Bayesian optimization.

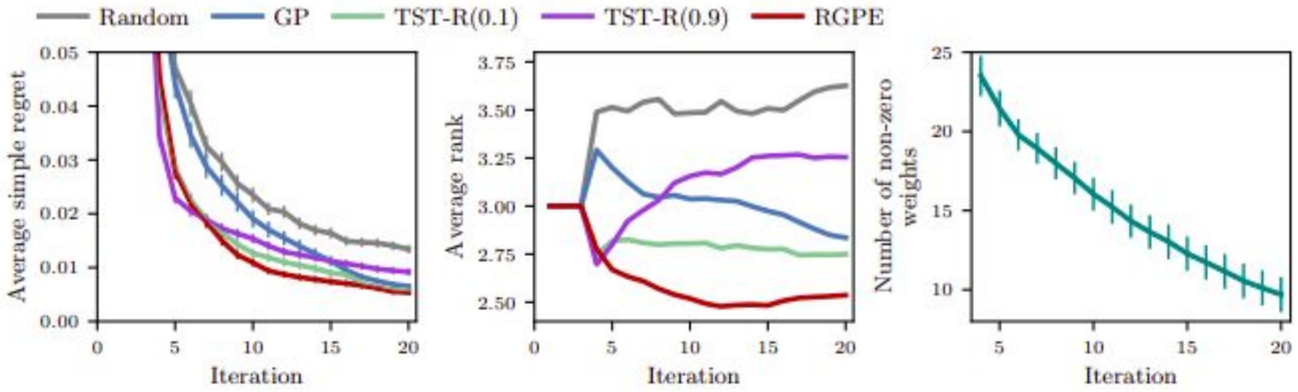


Figure 3: Optimization performance on the SVM hyperparameter optimization benchmarks, evaluated over 20 runs for each of 50 problems. (Left) Simple regret averaged over runs, with bars showing standard error. Warm-starting provided an initial boost over GP, and RGPE achieved the lowest regret. (Middle) The average rank of each method, ranked by simple regret, shows that RGPE consistently performed the best (lower is better). (Right) The number of non-zero weights in the RGPE ensemble. With 49 base models, more than half were immediately given zero weight, and by the end of the optimization only 10 models had positive weight.

Meta-Learning for Algorithm and Hyperparameter Optimization with Surrogate Model Ensembles

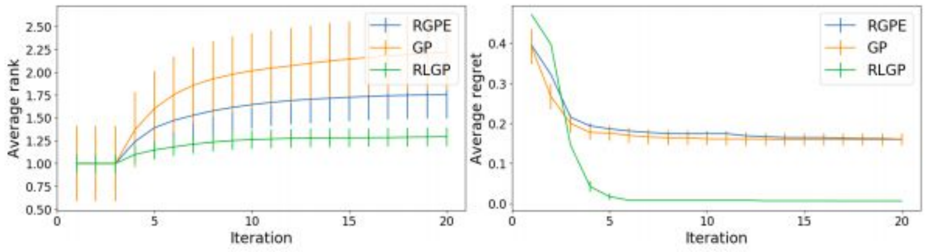


Figure 2: Evolution of average rank and regret using meta-data from 11 WEKA classification algorithms evaluated on 50 OpenML datasets from [3]. Error bars show ±1 standard deviation.

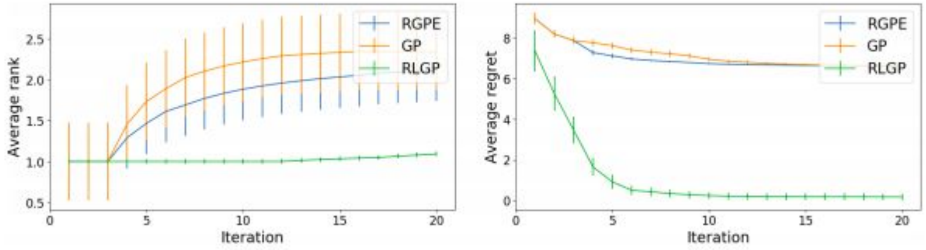


Figure 3: Results on synthetic datasets covering 6 global optimization functions from [6]

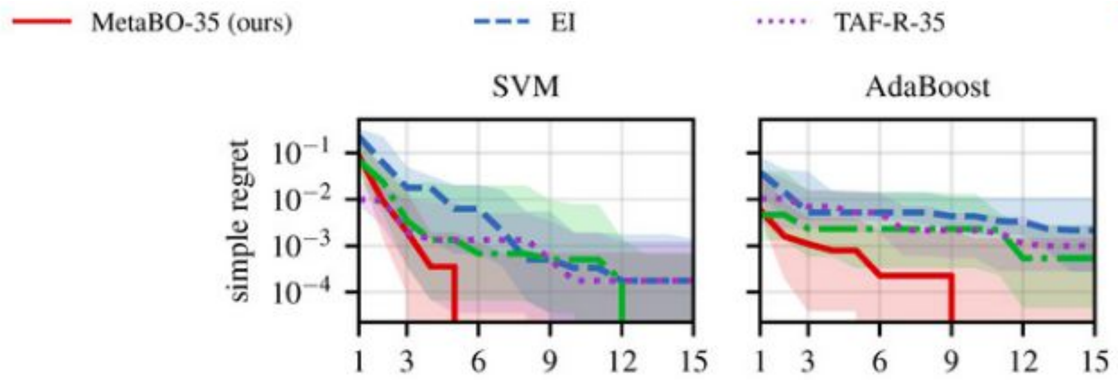
CASH problem

11 algorihtms

42000 hyperparameters conf.

50 randomly selected datasets OpenML

- Meta-Learning Acquisition Functions for Transfer Learning in Bayesian Optimization (ICLR 2020)



- *Transferable Neural Processes for Hyperparameter Optimization* (NIPS 2019)
- *Learning to Tune XGBoost with XGBoost* (NIPS 2019)

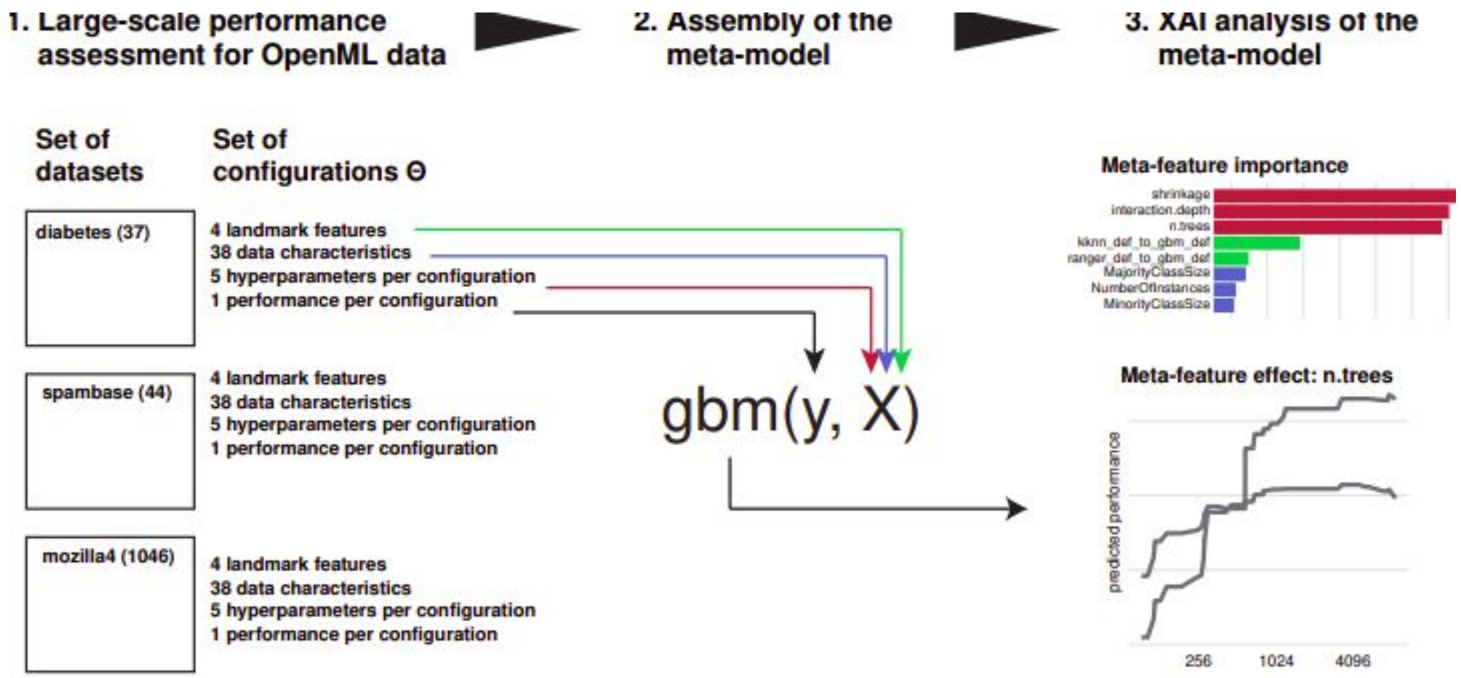
Towards explainable meta-learning

Katarzyna Woźnica[1] and Przemysław Biecek[1,2]

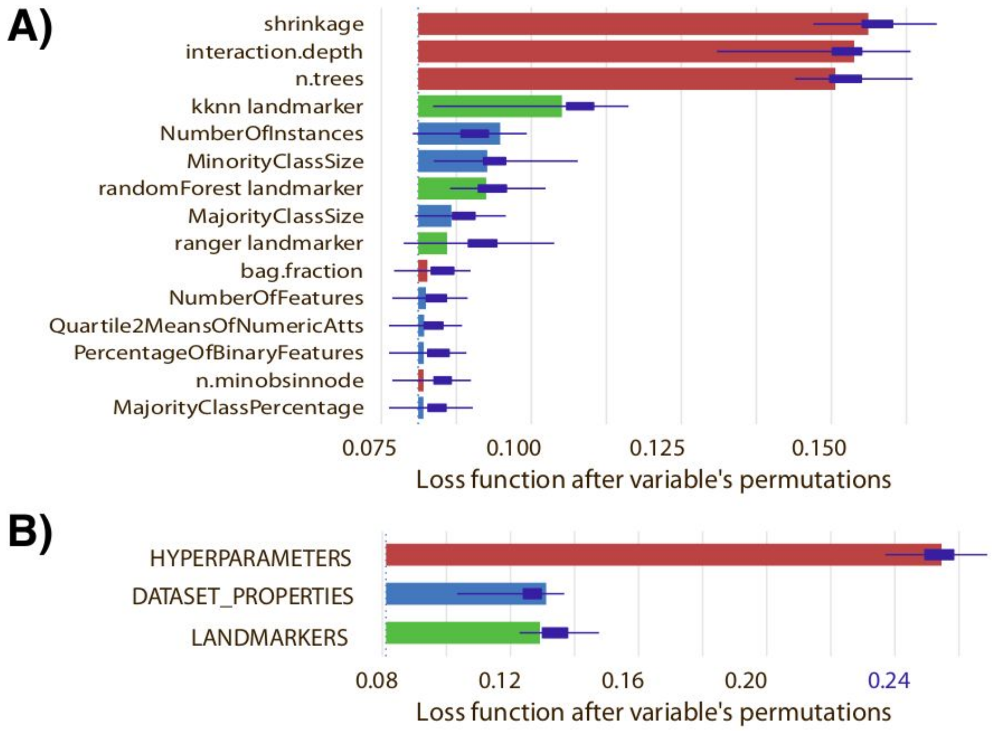
[1] Warsaw University of Technology
[2] Samsung R&D Institute Poland
katarzyna.woznica.dokt@pw.edu.pl

Abstract. Meta-learning is a field that aims at discovering how different machine learning algorithms perform on a wide range of predictive tasks. Such knowledge speeds up the hyperparameter tuning or feature engineering. With the use of surrogate models various aspects of the predictive task such as meta-features, landmarker models e.t.c. are used to predict the expected performance. State of the art approaches are focused on searching for the best meta-model but do not explain how these different aspects contribute to its performance. However, to build a new generation of meta-models we need a deeper understanding of the importance and effect of meta-features on the model tunability. In this paper, we propose techniques developed for eXplainable Artificial Intelligence (XAI) to examine and extract knowledge from surrogate models. To our knowledge, this is the first paper that shows how post-hoc explainability can be used to improve the meta-learning.

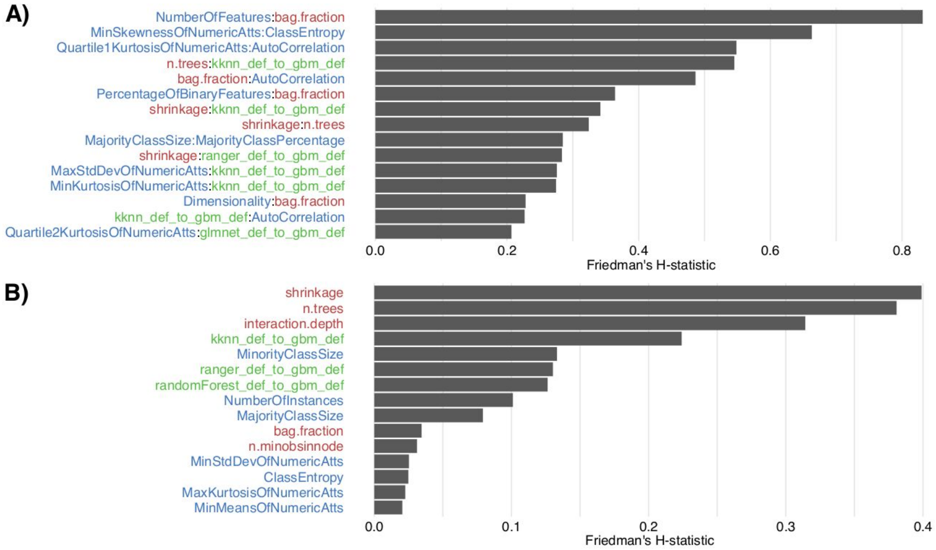
The Meta-OpenML100 surrogate model



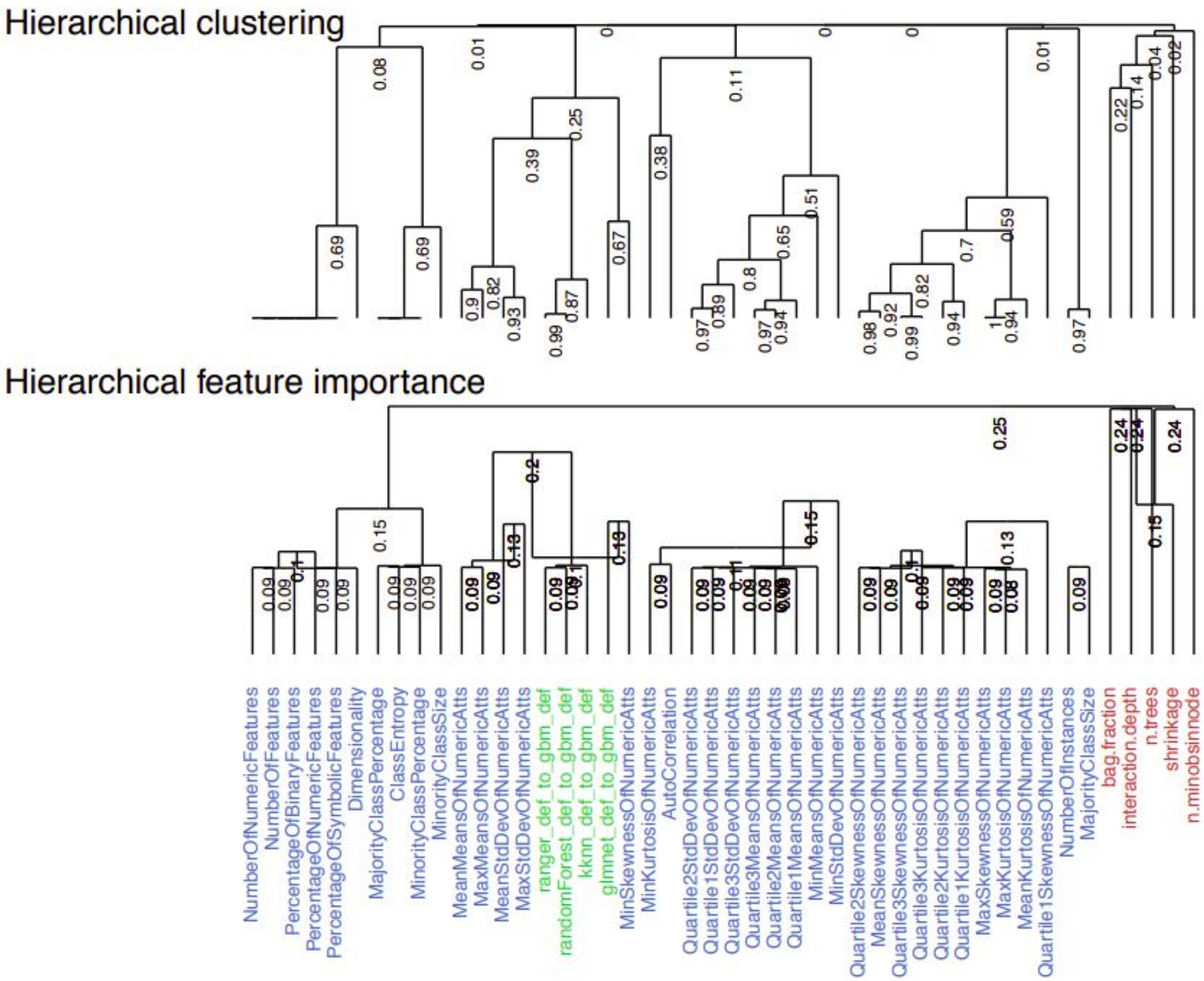
Meta-features importance



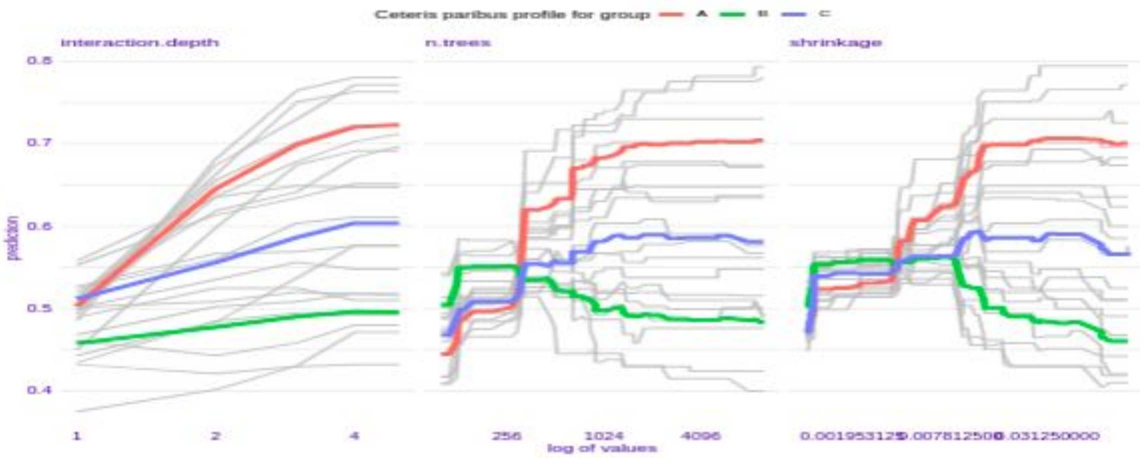
Meta-features interaction



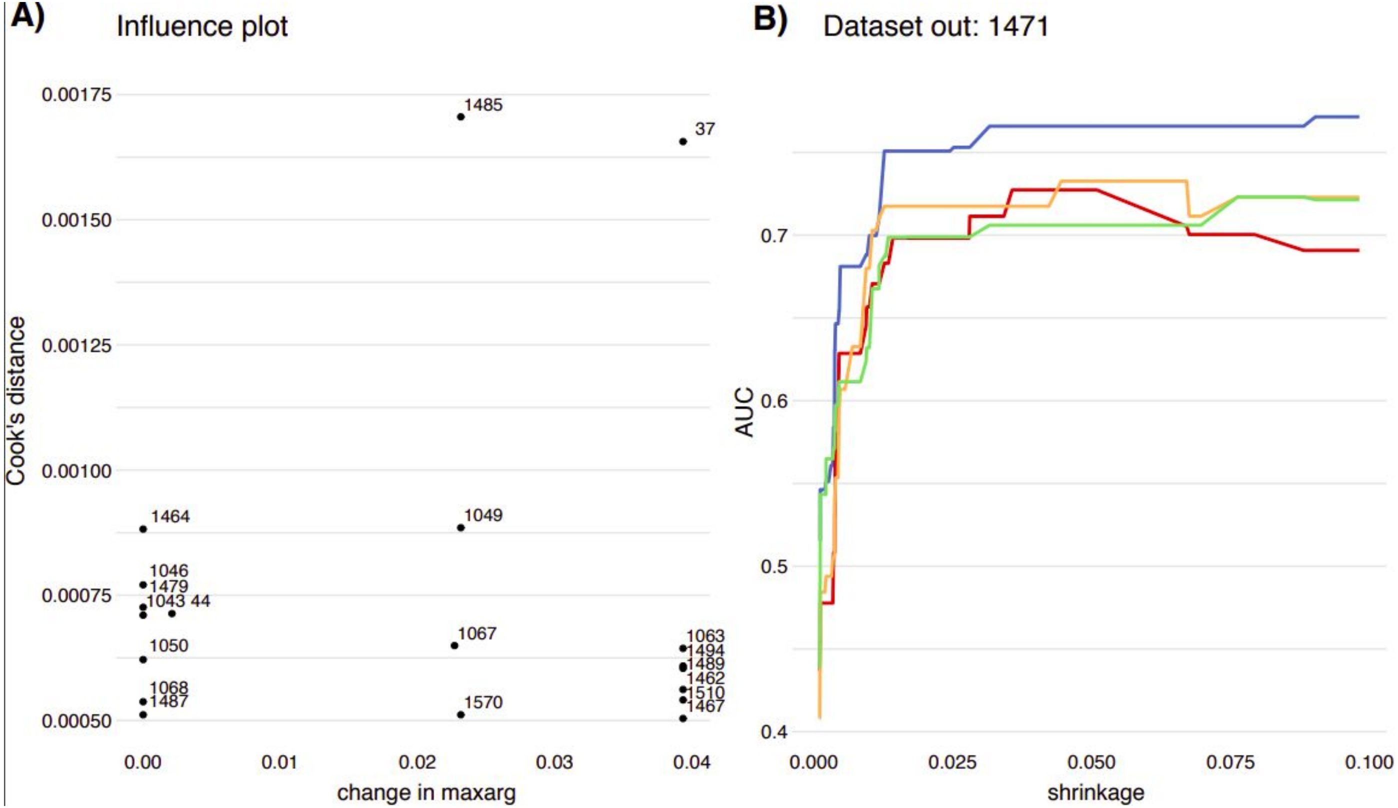
Importance of correlated meta-features



Hyperparameters informativeness



Robustness of meta-data



- XAI techniques are applied to extract knowledge about the importance of the particular meta-features in the meta-model
- This approach is universal and generic to the explainable analysis of any meta-learning model.

We may enhance the meta-model approach to better understanding meta-space.

Futher works:

- build new meta-model algorithm for boosting algorithms
- domain specific meta-learning (medical, credit data)