Proceedings of Machine Learning Research 191:1–7, 2022 ECMLPKDD Workshop on Meta-Knowledge Transfer

Advances in Metalearning: ECML/PKDD Workshop on Meta-Knowledge Transfer

Pavel Brazdil PBRAZDIL@INESCPORTO.PT
University of Porto, Portugal

Jan N. van Rijn JVRIJN@LIACS.NL
Leiden Institute of Advanced Computer Science, Leiden University, the Netherlands

Henry Gouk HENRY.GOUK@ED.AC.UK
University of Edinburgh, Scotland, United Kingdom

Felix Mohr FELIX.MOHR@UNISABANA.EDU.CO
Universidad de La Sabana, Colombia

Editors: P. Brazdil, J. N. van Rijn, H. Gouk and F. Mohr

Foreword

Meta-knowledge plays an important role in current machine learning and AutoML systems (Brazdil et al., 2022; Hutter et al., 2019). One way of acquiring meta-knowledge is by observing learning processes (on the same task, or different tasks) and representing it in such a way that it can be used later to improve future learning processes. Metalearning systems, on the other hand, normally explore metaknowledge acquired on different problems (Hospedales et al., 2021; Huisman et al., 2021; Vanschoren, 2018). The systems may, in addition, use meta-knowledge concerning which part of the space should be examined first (i.e., a warm start or dynamic scheduling).

Various contributions of this workshop addressed various aspects of metaknowledge, and in particular, how it is exploited in different systems. This workshop included two invited talks, one by Hospedales on "Meta-learning for Knowledge Transfer" and another by Hitzler (2022) on "Some advances regarding ontologies and neuro-symbolic artificial intelligence". The accepted papers can be divided into the following groups:

- Metalearning approaches for algorithm selection and hyperparameter optimization
- Metalearning approaches for cross-domain transfer
- Automatic methods for reformulating configuration spaces
- Instance selection in scheduling SAT and MIP problems

Although none of the papers addressed the issue of (meta-)knowledge transfer among different systems, we believe that this is an important issue to be addressed in future research (see Section 3). More details about both the invited talks and the accepted papers are presented in the following sections.

1. Invited Talks

The first invited talk entitled **"Meta-learning for Knowledge Transfer"** was given by Timothy Hospedales. This talk focused on recent work concerned with transferable and interpretable meta-learning in neural networks. The talk began with Hospedales providing pseudocode for a generic neural network training loop. Subsequent sections of his talk discussed how each of the pieces in this loop could be replaced with meta-learned components. It was demonstrated how one can learn new transferable loss functions that aid in learning from noisy labels (Gao et al., 2021) or learning in the presence of distribution shift (Gao et al., 2022b). Other parts of the talk described different approaches for initializing neural networks that achieve good performance when trained on very limited volumes of labelled data (Gouk et al., 2021; Hu et al., 2022; Zhang et al., 2021). Finally, the talk was concluded by some recent work from Hospedales's group on meta-learning new optimization algorithms for use with neural networks (Gao et al., 2022a), parameterized in such a way that the optimizers can be transferred to new machine learning problems and optimization experts can analyze the resulting methods to determine in which situations they can be expected to work well.

Hitzler (2022) presented an invited talk entitled **"Some advances regarding ontologies and neuro-symbolic artificial intelligence"**. The area of neural-symbolic artificial intelligence has attracted many researchers recently and resulted in many publications, including one book (Hitzler and Sarker, 2021). The area of neural-symbolic joins symbolic artificial intelligence (based on knowledge representation and logic) and artificial intelligence based on deep neural networks. The objective of the research is to propose methods that would explore the construction of systems that would reuse, adapt and integrate the methods and techniques from both areas and this way minimize the inherent limitations in each of these areas alone. As this is a vast area, Hitzler has focused in his invited talk on two issues. The first one covered Hitzler's work on concept induction over ontologies used for explaining deep learning systems. The other topic covered the acquisition of formal logical reasoning capabilities over ontologies through deep learning. More details can be found in the extended abstract, which includes various references to the works published on the topic.

2. Accepted Papers

The submitted papers went through a rigorous peer-review process, where each paper was assigned at least three reviewers. The reviewed papers were evaluated based on the feedback that the reviewers gave, rather than the aggregated review scores.

2.1. Metalearning approaches for algorithm selection and hyperparameter optimization

Gosiewska et al. (2022) contributed the extended abstract **"Interpretable Meta-Score for Model Performance"**, which summarises a paper that will appear in Nature Machine Intelligence. This paper proposes a novel approach to comparing learning algorithms that have been evaluated on multiple benchmark problems. The key idea is that of designing a model that is capable of answering the question of which method will perform best on a new problem. This is accomplished by fitting a logistic regression model that models the probability that one method will perform better than the other. Subsequently, one can interpret the coefficients associated with each learning algorithm as a meta-score for that approach. The full paper provides a rigorous comparison of several collections of benchmark datasets, including collections from OpenML and VTAB.

The paper **"Searching in the Forest for Local Bayesian Optimization"** by Deng and Lindauer (2022) explores recent findings that hyperparameter optimization landscapes are easier to optimize than formerly thought. They do so by employing a hierarchical two-stage procedure, dubbed BOingG, with in the first stage a Bayesian optimization procedure using a random forest is used to map the global structure of the hyperparameter optimization landscape, and in the second stage, the promising areas are being further searched using Bayesian optimization using Gaussian process techniques. They evaluate the approach on both several synthetic functions as well as several hyperparameter optimization problems involving ParamNet, and conclude that BOinG performs better, especially when more search budget is provided.

The paper **"Trust Region Meta Learning for Policy Optimization"** by Occorso et al. (2022) addresses a hyperparameter optimization problem for reinforcement learning algorithms. The authors argue that many reinforcement learning algorithms are being developed, all with many hyperparameters. They aim to solve this by training a meta-reinforcement learning agent, trust region policy optimization, that utilizes information from previous runs to become better at this task. The authors evaluate this on three reinforcement learning tasks, cartpole, mini-golf and Elikoeidis, and conclude that trust region policy optimization obtains better performance than the selected baselines.

2.2. Metalearning approaches for cross-domain transfer

The paper **"NeurIPS'22 Cross-Domain MetaDL Competition: Design and baseline results"** was contributed by Carrión-Ojeda et al. (2022). This paper outlines the design of a meta-learning competition accepted in the NeurIPS 2022 Competitions track. Most previous challenges in this area have focused on within-domain meta-learning, where training tasks and testing tasks are synthesised from the same dataset. In contrast, this competition uses different sources of data to generate training and testing episodes, resulting in a domain shift that reduces performance. This competition is facilitated by the construction of Meta-Album (Ullah et al., 2022), a meta-dataset that contains 40 different image classification datasets from 10 diverse domains. The competition has a strong focus on fostering an open-source meta-learning community by requiring code submissions and open-sourcing of the winning solutions.

The extended abstract entitled **"Experiments in Cross-domain Few-shot Learning for Image Classification"** by Wang et al. (2022) is based on a full version of the paper published in a journal. As the title suggests, it discusses a methodology and some results in the area known under the name Cross-domain Few-shot Learning. The authors evaluate the effect of using features extracted by different ResNet backbones at various levels of their convolutional hierarchies. It appears that a somewhat simple variant ResNet101-conv4 appears to lead to better results overall. The features extracted in this manner were used in conjunction with about ten different classifiers, although the extended abstract shows the results with LDA, 1-vs-rest logistic regression with l2 regularization and cosine similarity classifier. The experiments were carried out on five different domains and show that 1-vs-rest logistic regression with l2 regularization performs well overall.

2.3. Automatic methods for reformulating configuration spaces

The extended abstract **"On Usefulness of Outlier Elimination in Classification Tasks"** by Hetlerovič et al. (2022) is a summary of the full paper presented at the Intelligent Data Analysis Symposium (IDA 2022). It sheds light on the effect and importance of some outlier elimination in classification pipelines (workflows). The objective is to identify the most useful outlier elimination methods. To achieve this aim, the approach proceeds in two phases. In the first phase, it identifies the top-performing workflows, for each dataset, based on a metric A3R that combines accuracy and runtime. This is followed by the elimination of workflows that include infrequent outlier elimination methods. The authors have shown that the inclusion of these outlier elimination methods is indeed useful, as it leads to a lower loss in accuracy when different workflows are tested. There are other benefits as well. The reduced portfolio includes less than half of the original workflow. The method has identified just 3 outlier elimination methods (from the initial set of 12) that are generally useful for the given set of classification tasks on 50 datasets from OpenML (Vanschoren et al., 2014).

The paper **"Faster Performance Estimation for NAS with Embedding Proximity Score"** by Franken et al. (2022) addressed the problem of performance approximation in Neural Architecture Search (NAS). The work is closely related to an approach based on embeddings presented by Luo et al. (2018). The idea is to learn the performance of neural architectures in an embedded space in which gradient descent can be executed to identify new network architecture by decoding them from the embedding. The main contribution is to use a kNN approach in the embedding space instead of learning the surface with a neural network itself. While the approach cannot exploit gradients in the surface, it can quickly predict performance estimates for predefined candidate architectures. Experiments on the CIFAR10 dataset suggest that the method achieves state-of-the-art performance prediction quality while being significantly faster than existing methods.

2.4. Instance selection in scheduling SAT and MIP problems

The paper **"Instance selection for configuration performance comparison"** by Anastacio et al. (2022) addressed the problem of searching for the potentially best algorithm configuration of parameters. This study was oriented to SAT and MIP problems, each of which involves two datasets with many instances. The authors have noticed that time is often wasted on less promising configurations which often require a long running time regardless of the configuration. The method proposed exploits this knowledge. The authors have adapted four instance selection methods from the literature to work with the performance model used in model-based configurators. The methods exploit various measures characterizing their running time distribution. The results suggest that a decision can be reached 5 to 3000 times faster than with random sampling.

3. Future Vision

In this section, we present some reflections on the role of metaknowledge and its relationship to configuration spaces.

In recent years researchers have investigated various issues that are useful for the definition of configuration spaces. Some studies have addressed the problems of which hyperparameters are important, others on which features should be considered in the process of learning, or how these should be organised (e.g., proposals concerning the hierarchy of features). It is interesting to note that the information concerning configuration spaces can be considered as metaknowledge. Currently, the set-up of configuration spaces is normally done by the user. As such, this meta-knowledge comes typically from humans as a result of their prior experience in the area of machine learning. Despite the fact that some work was done in this direction, it is important to put more effort into methods that would enable to facilitate this task.

Deep neural networks have attracted a lot of attention, as these have proved useful in solving various tasks. Researchers working in this area usually point out that their advantage is that it is not necessary to worry about feature definition, nor selection, as this is done automatically by the system. But this has also disadvantages, however, as it is difficult to examine how the system arrived at a particular solution or decision (i.e. lack of explainability). More importantly, it is not easy to interact with these systems to correct a certain line of reasoning. It would seem that if complex processes were separated into a set of smaller tasks, this would increase the possibility of this interaction. The results at each stage could be represented in a symbolic form, which is typically used also to represent meta-knowledge and also the user's or other system's inputs. Future research could address this issue.

One important issue is how to transfer the (meta-)knowledge among different machine learning and AutoML systems so that their joint capability to solve problems would be enhanced. The individual AutoML systems could, in this context, be considered as 'agents' that should be endowed with the capability of transferring/communication this meta-knowledge to other systems, enhancing thus their joint capacity to solve more diverse and more complex problems.

References

Marie Anastacio, Théo Matricon, and Holger Hoos. Challenges of acquiring compositional inductive biases via meta-learning. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Pavel Brazdil, Jan N. van Rijn, Carlos Soares, and Joaquin Vanschoren. *Metalearning: Applications to Automated Machine Learning and Data Mining.* Springer, 2nd edition, 2022.

Dustin Carrión-Ojeda, Hong Chen, Adrian El Baz, Segio Escalera, Chaoyu Guan, Isabelle Guyon, Ihsan Ullah, Xin Wang, and Wenwu Zhu. Neurips'22 cross-domain metadl competition: Design and baseline results. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Difan Deng and Marius Lindauer. Searching in the forest for local bayesian optimization. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Gideon Franken, Prabhant Singh, and Joaquin Vanschoren. Faster performance estimation for nas with embedding proximity score. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Boyan Gao, Henry Gouk, and Timothy M Hospedales. Searching for robustness: Loss learning for noisy classification tasks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6670–6679, 2021.

Boyan Gao, Henry Gouk, Hae Beom Lee, and Timothy M Hospedales. Meta mirror descent: Optimiser learning for fast convergence. *arXiv preprint arXiv:2203.02711*, 2022a.

Boyan Gao, Henry Gouk, Yongxin Yang, and Timothy Hospedales. Loss function learning for domain generalization by implicit gradient. In *International Conference on Machine Learning*, pages 7002–7016. PMLR, 2022b.

Alicja Gosiewska, Katarzyna Woźnica, and Przemysław Biecek. Interpretable meta-score for model performance: Extended abstract. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Henry Gouk, Timothy M Hospedales, and Massimiliano Pontil. Distance-based regularisation of deep networks for fine-tuning. In *International Conference on Learning Representations*, 2021.

Dušan Hetlerovič, Luboš Popelínský, Pavel Brazdil, Carlos Soares, and Fernando Freitas. On usefulness of outlier elimination in classification tasks: Extended abstract. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Pascal Hitzler. Some advances regarding ontologies and neuro-symbolic artificial intelligence. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Pascal Hitzler and Md. Kamruzzaman Sarker, editors. *Neuro-Symbolic Artificial Intelligence: The State of the Art*, volume 342 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2021.

Timothy M Hospedales, Antreas Antoniou, Paul Micaelli, and Amos J. Storkey. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M Hospedales. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9068–9077, 2022.

Mike Huisman, Jan N van Rijn, and Aske Plaat. A survey of deep meta-learning. *Artificial Intelligence Review*, 2021.

Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren, editors. *Automated Machine Learning: Methods, Systems, Challenges*. The Springer Series on Challenges in Machine Learning. Springer, Cham, 1 edition, 2019.

Renqian Luo, Fei Tian, Tao Qin, Enhong Chen, and Tie-Yan Liu. Neural architecture optimization. In *Advances in Neural Information Processing Systems 31*, pages 7827–7838, 2018.

Manuel Occorso, Luca Sabbioni, Alberto Maria Metelli, and Marcello Restelli. Trust region meta learning for policy optimization. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Ihsan Ullah, Dustin Carrion, Sergio Escalera, Isabelle Guyon, Mike Huisman, Felix Mohr, Jan N. van Rijn, Haozhe Sun, Joaquin Vanschoren, and Phan Anh Vu. Meta-album: Multi-domain meta-dataset for few-shot image classification. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, 2022.

Joaquin Vanschoren. Meta-Learning: A Survey. *arXiv preprint arXiv:1810.03548*, 2018.

Joaquin Vanschoren, Jan N. van Rijn, Bernd Bischl, and Luis Torgo. OpenML: Networked science in machine learning. *ACM SIGKDD Explorations*, 15:49–60, 2014.

Hongyu Wang, Huon Fraser, Henry Gouk, Eibe Frank, Bernhard Pfahringer, Michael Mayo, and Geoff Holmes. Experiments in cross-domain few-shot learning for image classification: Extended abstract. In *ECMLPKDD Workshop on Meta-Knowledge Transfer*, volume 191 of *PMLR*, 2022.

Xueting Zhang, Debin Meng, Henry Gouk, and Timothy M Hospedales. Shallow bayesian meta learning for real-world few-shot recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 651–660, 2021.