From Automated to On-The-Fly Machine Learning

Presentation of work originally published in Machine Learning 107

Felix Mohr[1], Marcel Wever[1], Alexander Tornede[1], Eyke Hüllermeier[1]

Keywords: automated machine learning; classification; on-the-fly computing

Automated machine learning (AutoML) is the task of automatically selecting and parametrizing machine learning algorithms, as well as combining them into an overall solution (a "machine learning pipeline") specifically tailored for a task at hand (typically specified by a dataset). Existing approaches to AutoML are based on Bayesian optimization (e.g. auto-sklearn [Fe15]) or genetic algorithms (e.g. TPOT [Ol16]).

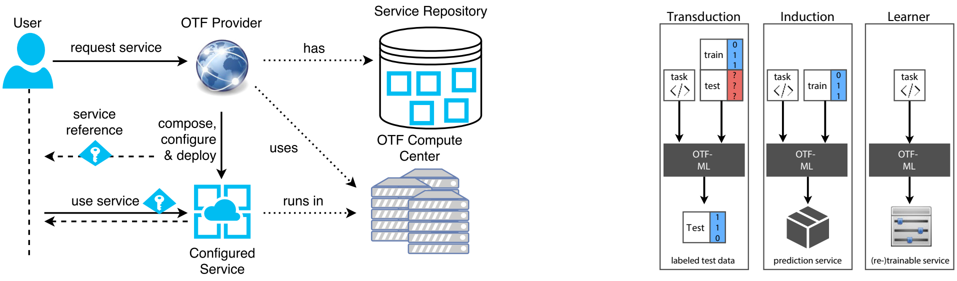
We recently complemented the repertoire of state-of-the-art AutoML tools by ML-Plan [MWH18b, WMH18, We19]. ML-Plan leverages techniques from hierarchical task network (HTN) planning to arrange the more than 10^{40} different candidate pipelines in a tree-shaped search space. In an extensive series of experiments, we showed that ML-Plan is highly competitive and often outperforms existing approaches.

Building on ML-Plan, our current work is devoted to the vision of what we call "On-the-Fly Machine Learning" (OTF-ML) — an instantiation of the On-the-Fly (OTF) computing paradigm [Ha13] for the case of machine learning, and, as such, an extension of AutoML. OTF computing aims at the provision of individually configured IT services in a dynamic, distributed market environment, which comprises different types of agents and allows customers to request services specifically tailored for their needs (cf. Fig. 1a).

In OTF-ML, we distinguish three types of services a customer may request (cf. Fig. 1b). In the *Transduction* scenario, the customer is interested in automatically labeling data. To this end, he provides a task description, along with training data and the data to be labeled. Internally, the OTF provider configures an ML service with the help of the provided training data and returns the labels for the unlabeled data obtained by the ML service. In the *Induction* scenario, the request only specifies the task and the training data. The customer is then provided access to the configured ML service, which can be queried to make predictions for new data points. Lastly, in the *Learner* scenario, the customer only describes the type of ML problem to be solved. He then obtains an ML service specifically tailored for such problems, which can be used for learning on whatsoever training data.

[1] Paderborn University, Warburger Straße 100, 33098 Paderborn, Germany <firstname>.<lastname>@upb.de

doi:10.18420/inf2019_40



a) Process of OTF provision of customized services

b) On-the-fly ML scenarios

Fig. 1: OTF-ML: the on-the-fly selection, configuration, provision, and execution of machine learning and data analytics functionality as requested by an end-user.

Realizing automated machine learning in an OTF environment offers various opportunities, including better computational resources, high parallelization and the combination of algorithms implemented for different platforms. In [Mo18, MWH18a], we extended ML-Plan to work on a service level combining implementations across plattforms. We consider these as important first steps paving the way for OTF-ML.

References

[Fe15] Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J. T.; Blum, M.; Hutter, F.: Efficient and Robust Automated Machine Learning. In: Advances in Neural Information Processing Systems 28. 2015.

[Ha13] Happe, M.; Meyer auf der Heide, F.; Kling, P.; Platzner, M.; Plessl, C.: On-The-Fly Computing: A Novel Paradigm for Individualized IT Services. In: 16th Int. Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing. 2013.

[Mo18] Mohr, F. Wever, M. Hüllermeier, E. Faez, A. Towards the Automated Composition of Machine Learning Services. In: IEEE International Conference on Services Computing. 2018.

[MWH18a] Mohr, F.; Wever, M.; Hüllermeier, E.: Automated Machine Learning Service Composition. CoRR, abs/1809.00486, 2018.

[MWH18b] Mohr, F.; Wever, M.; Hüllermeier, E.: ML-Plan: Automated Machine Learning via Hierarchical Planning. Machine Learning, 107(8-10):1495–1515, 2018.

[Ol16] Olson, R. S.; Urbanowicz, R. J.; Andrews, P. C.; Lavender, N. A.; Kidd, L. C.; Moore, J. H.: Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. Applications of Evolutionary Computation, 2016.

[We19] Wever, M.; Mohr, F.; Tornede, A.; Hüllermeier, E.: Automating Multi-Label Classification Extending ML-Plan. 6th ICML Workshop on Automated Machine Learning, 2019.

[WMH18] Wever, M.; Mohr, F.; Hüllermeier, E.: ML-Plan for Unlimited-Length Pipelines. 5th ICML Workshop on Automated Machine Learning, 2018.