Federated Meta-Learning: Democratizing Algorithm Selection Across Disciplines and Software Libraries

-- Proposal, v1 --

Joeran Beel
Trinity College Dublin
School of Computer Science and Statistics
Artificial Intelligence Discipline
ADAPT Centre
beelj@tcd.ie

KEYWORDS

Introduction
There is an ever-growing number of tools for automating the machine learning pipeline, both commercial and open source. Auto-sklearn [11, 15], Auto Weka [14], ML-Plan [18], and H2O.ai are only some examples. In addition, the auto-movement reached other areas of data science in the wider sense, for instance, the recommender-system community with LibRec-Auto [17]. A key component many of these auto- and algorithm selection tools is the algorithm selection and configuration process, which is subject to intensive research [1–3, 6, 7, 9, 10, 12, 13, 16, 20, 22, 24]. In some disciplines, like the SAT community, automated algorithm selection tools like SATzilla achieved remarkable improvements compared to ‘standard’ algorithms [25].

Meta-learning is one of the most promising techniques to warm-start the algorithm selection and configuration process [12]. With meta-learning, a machine learning model is trained to predict how algorithms will perform on a given task. That predicted algorithm can then either be applied directly to solve the task, or further be optimized. The meta-learning model is built based on the past performance of algorithms on a large number of tasks (datasets). The tasks are typically described through meta-features, and for new unseen tasks, the most performant algorithms can be predicted through the meta-learner. Meta-learning is not only used for predicting the performance of machine learning algorithms but for almost any kind of algorithm selection in various disciplines (SAT [25]; recommender systems [5, 8]; reference parsing [21]...).

Research Problem
The prediction accuracy of meta learning varies strongly among disciplines. A challenge is the (non) availability of data in some disciplines to build the meta-learning model. This is due to the typical workflow of machine learning, data science etc. Typically, software libraries – be it machine learning libraries like (Auto) sklearn, or recommender system libraries like LibRec (-Auto) are used in isolation, either locally or in the cloud. Either way, the information how different algorithms and their configurations performed on the particular dataset, is neither published nor shared with others.

Research Goal
Our goal is to facilitate the algorithm selection process by leveraging historic performance data that was on various devices by various software libraries.

Federated Meta Learning
We propose “Federated Meta-Learning” (FML), a concept that allows everyone to benefit from the data that is generated through software libraries including standard machine learning and data science libraries as well as the auto tools. We envision a peer-to-peer or client-server architecture that allows the exchange of meta-data and models for the purpose of meta-learned algorithm selection and configuration.

The input to Federated Meta Learning is a description of the task, and the output is a recommendation for the potentially best performing algorithm(s) to solve that task. This recommendation could consist simply of a list of the best algorithms, or their predicted performance values. The list could also consist of multiple sub-lists created with different meta learners.

In its simplest form, federated meta-learning would simply be a knowledge base or directory of algorithms/data performance measures. Ultimately, Federated Meta-Learning would be able to predict algorithm performance for unseen tasks.

To the best of our knowledge this concept is novel. The term “Federated Meta Learning” has only been used once before by Chen et al. in a different context [4]. Federated Machine Learning has – unlike distributed machine learning [19] or central repositories like OpenML [23] – no central instance that stores and controls all (raw) data. Instead, the learning is performed on the local devices who remain the data owners and in full control. This makes Federated Machine Learning similar to the recently introduced concept of “federated machine learning” by Google: “Federated [Machine] Learning enables [devices] to collaboratively learn a shared prediction model while keeping all the training data...
on device, decoupling the ability to do machine learning from the need to store the data in the cloud. ¹

However, Federated Machine Learning focuses on learning one machine learning task across multiple devices. Federated Meta Learning focuses on learning algorithm performance for arbitrary tasks across devices (or one generic model applicable to all tasks). We envision federated meta learning as an ecosystem where the raw data is kept on the original devices. However, either a) the meta data would be shared among the devices, b) the meta data would be stored on a central device, or c) the learning would be distributed, and the created model would be shared among the devices.

Challenges
Federated Meta Learning is a highly challenging research project. Given that Federated Meta Learning is a specialization of Federated Machine Learning, Federated Meta Learning faces all the challenges that exist in Federated Machine Learning, plus additional challenges. The key challenge lies in the question of how to effectively train models across the devices? Other challenges include the creation of a generic data description language that is powerful enough to satisfy the needs of different disciplines. Architectural challenges include the question whether to use a peer-to-peer or client-server architecture. In the long run, social questions need consideration such as preventing manipulation (developers of algorithms may have an interest that their algorithms are ‘recommended’) and free-rider problems (users benefiting from the system without sharing their data). Another challenge would arise if the system should not only focus on the globally best algorithm for a task (an entire dataset) but if per-instance algorithm selection should be learned. This would make the whole system even more complex.

REFERENCES


[20] Romero, C., Olmo, J.L. and Ventura, S. 2013. A meta-learning approach for recommending a subset of white-

¹ https://ai.googleblog.com/2017/04/federated-learning-collaborative.html
box classification algorithms for Moodle datasets. 


