

Federated Meta-Learning: A Novel Approach for Algorithm Selection

by Mukesh Arambakam - 19301497

Abstract

“Federated Meta-Learning” (FML), a concept that allows everyone to benefit from the data that is generated through software libraries including machine learning and data science libraries. It focuses on learning the algorithm performance measures and making recommendations based on the model created. I introduce FMLearn, an application developed using the client-server model, which allows the exchange of meta-data about machine learning models and data in itself, for the purpose of meta-learned algorithm selection and configuration. The input to FMLearn is a dataset and the output is a recommendation for the potentially best performing algorithm(s) and its hyper-parameters to solve the task. This recommendation is made by a model built using the Meta-Features obtained from the dataset description, using the K-Nearest Neighbours algorithm and historic performance data. Scikit-Learn’s toy datasets along with other datasets from UCI Machine Learning repository were used and evaluated against various machine learning algorithms using GridSearchCV and Cross-Validation for which the execution time was measured. In the case previously seen datasets like scikit-learn’s breast cancer dataset, an execution time of approximately 94.24min was recorded to find the best algorithm by performing Grid Search and Cross-Validation. Whereas, when FMLearn was asked to recommend the best algorithm, the application only took 3secs to recommend the best performing algorithm along with its model parameters. For a large dataset like the skin segmentation, traditional means for finding the best performing algorithms takes approximately 869.74 minutes and when FMLearn was asked to recommend the same, it took only 3secs. In the case of previously unseen but similar datasets, the recommended algorithm is accurate, but the hyper-parameters required re-optimisation to suite the dataset. For a previously unseen and highly dissimilar dataset, FMLearn recommends a list of algorithms which it thinks are best suited for that dataset based on its prior knowledge. In this case, FMLearn recommends the best performing algorithm about 60% of the time. Overall, the use of this application allows the user to scale down an average of 86.718% and 95.762% of time and electricity for small and large datasets respectively by eliminating the repetitive and time consuming task of algorithm selection and configuration from the Machine Learning Workflow.

Keywords

Federated Meta-Learning, FMLearn, AutoML, Machine Learning, RecSys, Dataset Meta-Features, Algorithm Selection, Algorithm Configuration