BOML: A MODULARIZED BILEVEL OPTIMIZATION LIBRARY IN PYTHON FOR META LEARNING

Yaohua Liu[†], *Risheng Liu*[‡,§,*]

[†]International School of Information Science & Engineering, Dalian University of Technology
[‡]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province
[§]Pazhou Lab, Guangzhou
liuyaohua_918@163.com, rsliu@dlut.edu.cn

2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) | 978-1-6654-4989-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICMEW53276.2021.9455948

ABSTRACT

Meta-learning (a.k.a. learning to learn) has recently emerged as a promising paradigm for a variety of applications. There are now many meta-learning methods, each focusing on different modeling aspects of base and meta learners, but all can be (re)formulated as specific bilevel optimization problems. This work presents BOML, a modularized optimization library that unifies several meta-learning algorithms into a common bilevel optimization framework. It provides a hierarchical optimization pipeline together with a variety of iteration modules, which can be used to solve the mainstream categories of meta-learning methods, such as meta-feature-based and meta-initialization-based formulations. The library is written in Python and is available at https://github.com/dut-media-lab/BOML.

***Index Terms*—** Bilevel optimization, meta-learning, dynamical system, few-shot classification, Python.

1. INTRODUCTION

Meta-learning aims to deal with the problem of "learning to learn" and has recently emerged as a potential learning paradigm that can gain experience over previous tasks and generalize that experience to unseen tasks proficiently. The applications of meta-learning span from few-shot classification [1], and deep Reinforcement Learning (RL) [2], to Neural Architecture Search (NAS) [3]. However, it actually requires significant optimization expertise to design efficient algorithms to solve these meta-learning problems due to the complex learning paradigms.

In this work, by formulating meta-learning tasks from the bilevel optimization perspective, we establish a unified and modularized library, named BOML, for different categories of meta-learning approaches. Specifically, in BOML, we support two main categories of meta-learning paradigms, including meta-initialization-based [2] and meta-feature-based [4] and implement a variety of recently developed bilevel optimization techniques, such as Reverse Hyper-Gradient (RHG), Truncated RHG (TRHG), Meta-SGD, MT-net, WarpGrad, HOAG and Bilevel Descent Aggregation (BDA), for solving the meta-learning problems. Several first-order approximation schemes, including First-Order MAML (FMAML) and DARTS, are also integrated into our BOML.

The key features of BOML can be summarized as follows: It provides a unified bilevel optimization framework to address different categories of existing meta-learning paradigms, offers a modularized algorithmic structure to integrate a variety of optimization techniques, and is flexible and extensible for potential meta-learning applications. We implement continuous code integration with *Travis CI* and *Codecov* to obtain high code coverage (more than 98%). We also follow *PEP8* naming convention to guarantee the code consistency. The documentations are developed with *sphinx* and rendered using *Read the Docs*.

We would like to thank support from National Natural Science Foundation of China (Nos. 61922019 and 61672125), LiaoNing Revitalization Talents Program (XLYC1807088) and the Fundamental Research Funds for the Central Universities. *Corresponding author.

2. A UNIFIED BILEVEL OPTIMIZATION PARADIGM FOR META LEARNING

We first present a general bilevel optimization paradigm to unify different types of meta-learning approaches. Specifically, we define the meta dataset as $\mathcal{D} = \{\mathcal{D}^i\}_{i=1}^N$, where $\mathcal{D}^i = \mathcal{D}^i_{\mathtt{tr}} \cup \mathcal{D}^i_{\mathtt{val}}$ is linked to the i-th task and $\mathcal{D}^i_{\mathtt{tr}}$ and $\mathcal{D}^i_{\mathtt{val}}$ respectively denote the training and validation sets. We denote the parameters of the base-learner as $\mathbf{y}^i$ for the i-th task. Then the meta-learner can be thought of as a function that maps the dataset to the parameters of base-learner for new tasks, that is, $\mathbf{y}^i = \Psi(\mathbf{x}, \mathcal{D}^i)$, where $\mathbf{x}$ is the parameter of the meta-leaner shared across tasks. With the above notations, we can formulate the general purpose of meta-learning tasks as the following bilevel optimization model:

$$\begin{aligned} &\min_{\mathbf{x}} F(\mathbf{x}, \{\mathbf{y}^i\}_{i=1}^N), \quad s.t. \\ &\mathbf{y}^i \in \arg\min_{\mathbf{y}^i} f(\mathbf{x}, \mathbf{y}^i),\ i = 1, \cdots, N, \end{aligned} \tag{1}$$

where $f(\mathbf{x}, \mathbf{y}^i) = \ell(\mathbf{x}, \mathbf{y}^i, \mathcal{D}^i_{\mathtt{tr}})$ and $F(\mathbf{x}, \{\mathbf{y}^i\}_{i=1}^N) = 1/N \sum_{i=1}^N \ell(\mathbf{x}, \mathbf{y}^i, \mathcal{D}^i_{\mathtt{val}})$ are called the Lower-Level (LL) and Upper-Level (UL) objectives, respectively. By reformulating the optimization process of $\mathbf{y}^i$ (with fixed $\mathbf{x}$) as a dynamical system, that is, $\mathbf{y}^i_0 = \Psi_0(\mathbf{x}, \mathcal{D}^i)$ and $\mathbf{y}^i_t = \Psi_t(\mathbf{x}, \mathbf{y}^i_{t-1}, \mathcal{D}^i)$, the meta-learner can be established as $\Psi = \Psi_T \circ \Psi_{T-1} \circ \cdots \circ \Psi_0$. Based on the above construction, we can formulate different categories of meta-learning methods within the bilevel model in Eq. (1).

3. DESIGN AND FEATURES OF BOML

3.1. Optimization Process

We first illustrate the optimization process of BOML for meta-learning in Figure 1. It can be seen that BOML constructs two nested optimization subproblems (blue and green dashed rectangles), which are respectively related to the LL variable $\mathbf{y}$ and UL variable $\mathbf{x}$ in Eq. (1). For the LL subproblem (w.r.t. $\mathbf{y}$), we can establish the dynamical system (parameterized by fixed $\mathbf{x}$) by performing Gradient Descent (GD) on the LL objective (that is, f).

We also consider the recently proposed aggregation technique in BDA to integrate both the LL and UL objectives (that is, f and F) to generate the dynamical system. As for the UL subproblem, we actually consider Back-Propagation (BP) and Back-Propagation-Through-Time (BPTT) [5] to calculate the gradients (w.r.t. $\mathbf{x}$) respectively.

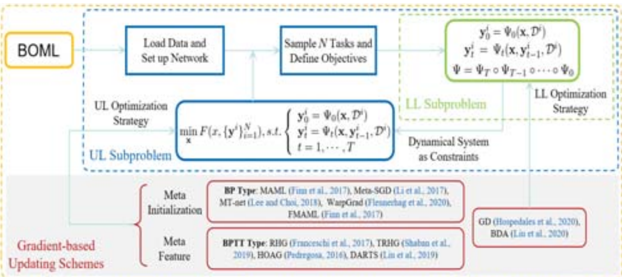

Fig. 1. Illustrating the optimization process of BOML.

3.2. Implementation Details

Here we consider few-shot classification [3] as our demo application and implement network structures for both meta-initialization-based and meta-feature-based approaches. We also provide loading configurations for several widely-used datasets, including Omniglot, MiniImageNet and so on. Users can easily build bilevel optimization model for meta-learning with brief code snippets.

3.3. Comparison to Existing Libraries

In the past few years, some libraries, such as Meta-Blocks [1] and Far-HO [6] [2], have also been developed for meta-learning. BOML establishes a general bilevel optimization framework to unify both above two categories of meta-learning methods, and implements a series of recently proposed extensions and accelerations in BOML. We also employ more recently proposed methods and new BLO techniques for gradient computation. More detail introduction for the comparison of related algorithms is provided in the online documentation.

4. CONCLUSIONS AND FUTURE WORKS

We (re)formulated mainstream meta-learning approaches into a unified bilevel optimization paradigm, and presented BOML to integrate common meta-learning approaches under our proposed taxonomy. For future work direction, we plan to extend more gradient computation modules and related methods to support RL stratigies and NAS.

5. REFERENCES

[1] Risheng Liu, Pan Mu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang, "A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton," in *International Conference on Machine Learning*, 2020, pp. 6305–6315.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, 2017, pp. 1126–1135.

[3] Risheng Liu, Jiaxin Gao, Jin Zhang, Deyu Meng, and Zhouchen Lin, "Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond," *arXiv preprint arXiv:2101.11517*, 2021.

[4] Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimilano Pontil, "Bilevel programming for hyperparameter optimization and meta-learning," *arXiv preprint arXiv:1806.04910*, 2018.

[5] Atılım Günes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind, "Automatic differentiation in machine learning: a survey," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5595–5637, 2017.

[6] Luca Franceschi, Michele Donini, Paolo Frasconi, and Massimiliano Pontil, "Forward and reverse gradient-based hyperparameter optimization," in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, 2017, pp. 1165–1173.

[1] Code: https://github.com/alshedivat/meta-blocks.
[2] Code: https://github.com/lucfra/FAR-HO.