Model-Agnostic Meta-Learning Techniques: A State-of-The-Art Short Review

Aikaterini I. Griva
ELEDIA@AUTh, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
aigriva@physics.gr

Achilles D. Boursianis
ELEDIA@AUTh, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
bachi@physics.auth.gr

Lazaros A. Iliadis
ELEDIA@AUTh, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
liliadis@physics.auth.gr

Panagiotis Sarigiannidis
Department of Electrical and Computer Engineering
University of Western Macedonia
Kozani, Greece
psarigiannidis@uowm.gr

George Karagiannidis
School of Electrical and Computer Engineering
Aristotle University of Thessaloniki
Thessaloniki, Greece
geokarag@auth.gr

Sotirios K. Goudos
ELEDIA@AUTh, School of Physics
Aristotle University of Thessaloniki
Thessaloniki, Greece
sgoudo@physics.auth.gr

2023 12th International Conference on Modern Circuits and Systems Technologies (MOCAST) | 979-8-3503-2107-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/MOCAST57943.2023.10176893

***Abstract*—In the last few years, novel meta-learning techniques have established a new field of research. Great emphasis is given to few-shot learning approaches, where the model is trained by using only a few training examples. In this work, a review of several model-agnostic meta-learning methodologies (MAML) is presented. Firstly, we identify and discuss the typical characteristics of the first proposed MAML algorithm. Next, we classify the model-agnostic approaches into three main categories: Regular gradient descent MAML, Hessian-free MAML, and Bayesian MALM, by presenting their advantages and limitations. Finally, we conclude this work with further discussion and highlight the future research directions.**

***Index Terms*—Model-agnostic meta-learning, meta-learning, deep learning, few-shot learning, one-shot learning**

I. Introduction

Humans have the great ability to learn quickly. We can recognize objects from a few examples, we can learn new tasks based on prior experience and a small amount of information, thus we are expecting our deep neural networks to perform the same. However, large amounts of data are required from such models to perform effectively. This is the reason why we are constantly investigating approaches such as few-shot learning techniques to decrease the amount of data and at the same time improve our models' performance.

Few-shot learning is a challenging approach to training a learning model by using only a few examples. For multiple tasks, a common feature representation can be learned using prior experience [1]. Recently, a meta-learning approach has been used to address the issue of few-shot learning. Meta-learning studies how intelligent systems can increase their performance through experience [2]. In other words, we can describe meta-learning as the mechanism for learning how to learn. In contrast to conventional approaches, meta-learning intends to enhance the learning algorithm itself, taking into account the knowledge gained from previous learning episodes [3]. One famous meta-learning method is the model-agnostic meta-learning (MAML) algorithm. MAML is characterized by its simplicity and the fact that can be applied to any problem whose solution is approximated with gradient descent.

The TERMINET project aims at providing a novel next-generation reference architecture based on cutting-edge technologies. Within this project, both the multi-task and meta-learning perspectives will be supported to enable personalized or device-specific modeling. The Meta-Learning technique will be used to optimize the performance of each device based on the given results of few-shot adaptation examples on heterogeneous tasks. Moreover, MAML algorithms will be considered and adopted as guiding approaches that offer improved personalized models for a large majority of the Internet of Things (IoT) end devices.

Motivated by the above, the current contribution provides a state-of-the-art review of the existing research from the area of MAML methodologies. In detail, the contribution of this paper is summarized as follows:

1) Three state-of-the-art MAML techniques, namely regular gradient descent MAML, hessian-free MAML, and bayesian MAML are identified and thoroughly studied.
2) The principles, advantages, and limitations of each algorithm are highlighted.

The rest of this paper is organized as follows: Section II introduces the first MAML algorithm that was proposed in 2017. Section III presents the most popular MAML techniques and classifies them according to their characteristics. Section IV explores the future directions, whereas Section V concludes this work and summarizes its findings.

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957406 (TERMINET).

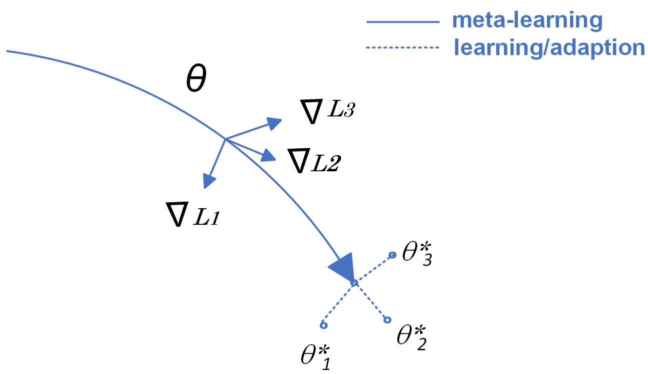

Fig. 1. Model-agnostic meta-learning algorithm representation

Algorithm 1 Pseudocode of the first proposed MAML algorithm

Require: distribution $p(\mathcal{T})$ over tasks
Require: hyperparameters step size α, β initialization
Initialize randomly the parameters w of the parameterized model f_w
while task **do**
  Sample tasks $\{\mathcal{T}_i\} \sim p(\mathcal{T})$
  for all $\{\mathcal{T}_i\}$ **do**
    Compute $\nabla_w \mathcal{L}_{\mathcal{T}_i}(f_w)$ w.r.t. N examples
    $w_i^{'} \leftarrow w - \alpha \nabla_w \mathcal{L}_{\mathcal{T}_i}(f_w)$
  end for
  $w \leftarrow w - \beta \nabla_w \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{w_i^{'}})$
end while

II. Model-Agnostic Meta-Learning algorithm

MAML is an algorithm for meta-learning that was proposed in 2017 by Chelsea Finn et al. [3]. It is called model agnostic because it was designed to perform with any model trained with gradient descent and also to be applicable to a wide range of learning problems mainly in the deep learning framework. The core idea is to train the initial parameters of the model to maximize its performance on a new task after updating those parameters, Fig. 1. This can be achieved through one or more gradient steps using a small amount of data from the new task. The proposed algorithm neither increases the number of learned parameters nor imposes restrictions on the model's architecture. Furthermore, it can be used with a variety of loss functions, and the idea is to maximize the sensitivity of those loss functions of new tasks with respect to the parameters.

Training the parameters of the model such that a few or even a single gradient step can lead to better results on a new task can be generalized as building an internal representation that is broadly applicable to a wide range of tasks. Finn et al. [3] experimentally evaluated the algorithm in few-shot classification, supervised regression, and reinforcement learning (RL) problems. However, the MAML algorithm has a number of issues that can limit the generalization performance of the model, reduce the flexibility of the framework, and lead to instabilities during training. Moreover, MAML can be computationally expensive during training and inference and requires the model to go through a costly hyperparameter tuning. In order to face those weaknesses, an improved subvariant of the MAML framework called MAML++ was proposed by Antreas Antoniou et al. in 2019 [4] and was experimentally evaluated in few-shot learning tasks achieving better results.

III. Popular MAML algorithm subvariants

During the last few years, several different subvariants of the classical MAML algorithm can be found in the literature. In this work, some popular techniques are briefly presented. To maintain consistency, the algorithms are classified into the following three categories:

- Regular gradient descent MAML algorithms,
- Hessian-free MAML algorithms, and
- Bayesian MAML algorithms.

A. Regular gradient descent MAML algorithms

The main characteristic of the methods that can be found in the first technique is that higher-order derivatives are used in the backpropagation of the error procedure. The algorithms are proposed to optimize the main MAML algorithm in terms of computational efficiency and stability. Moreover, they can be extended to problems with previously unseen classes.

- Implicit Model-Agnostic Meta-Learning (iMAML) was proposed by Aravind Rajeswaran et al. [5] in 2019. This algorithm was developed in order to achieve significant gains in computing and memory efficiency. iMAML gives the opportunity to select the suitable inner optimization method. In this algorithm, there is no need of differentiating through the optimization path. The algorithm's goal is to learn a set of parameters and lead to good generalization for various tasks based only on the solution to the inner-level optimization. At the same time an iMAML subvariant, i.e. the Hessian-free subvariant was also proposed achieving slightly better results in the few-shot classification problem.
- Adaptive Model-Agnostic Meta-Learning (Alpha-MAML) was proposed by Harkirat Singh Behl et al. [6] in 2019 as an extension to MAML. Alpha-MAML increases the algorithm's resistance to hyperparameter choices and improves stability. This algorithm was evaluated on a few-shot image classification problem and the results showed that is more robust than MAML due to the less time that is required for tuning.
- Out-of-distribution (OOD-MAML) is another extension of the MAML algorithm, introduced by Taewon Jeong et al. [7] in 2020. Fake or out-of-distribution samples that look like in-distribution samples are used to train the model in order to detect those OOD samples in new tasks. This method achieves learning both the initialization of a model and the initial OOD-samples over the tasks. The experimental evaluation in classification tasks shows that

the proposed algorithm achieves higher accuracy than MAML and, at the same time, detects OOD samples.

B. Hessian-free MAML algorithms

Another interesting approach is the Hessian-free MAML methods that ignore second-order derivatives when backpropagating in the meta-learning procedure to save a huge amount in the network computation and at the same time maintain the performance of the model.

- First Order Model-Agnostic Meta-Learning (FOMAML) [3], [8] is a subvariant of the first proposed MAML that ignores second derivatives. In that way, it is simpler to implement and reduces computational complexity. However, it might lose gradient information. It is proven that FOMAML works as well as the MAML algorithm in few-shot classification problems without affecting its performance.
- Reptile is a method introduced by Alex Nichol et al. [8] in 2018. Reptile is similar to FOMAML and quite simple. It is characterized by its ability to learn an initialization for the model's parameters. After the optimization of these parameters, the model can be generalized from only a few examples. By evaluating the results of the algorithms in few-shot classification tasks, we obtain that they perform as well as the main MAML algorithm.
- Hessian-Free (HF-MAML) was proposed by Alireza Fallah et al. [9] in 2020 to achieve low computational complexity and, at the same time, to preserve all theoretical guarantees of the basic algorithm without evaluating any Hessian. The authors provide a theoretical analysis of HF-MAML and highlight its improvement regarding the convergence properties compared to MAML and FOMAML algorithms.

C. Bayesian MAML algorithms

The Bayesian hierarchical model is also used in meta-learning techniques to incorporate uncertainty into the model estimation. Based on this, there are several subvariants of the basic MAML algorithm that improve its performance metrics and weaknesses.

- Laplace Approximation for Meta-Adaption (LLAMA) is an extension to MAML based on the hierarchical Bayesian model and the Laplace approximation that provides a more accurate estimate of the integral. It was proposed by Erin Grant et al. [10] in 2018. The experimental evaluation of the one-shot classification problem shows a small improvement in accuracy. However, skewed distributions do not respond well to this approximation. Furthermore, the extension has limitations since it uses a point estimate to approximate the predictive distribution over additional data points. To overcome these limitations, the BMAML was proposed.
- Bayesian MAML (BMAML) was proposed in 2018 by Taesup Kim et al [11], in order to learn complex uncertainty structures and combine an effective nonparametric variational inference method with gradient-based meta-learning. This variant is effective, precise, and robust. At the same time, it is simple to implement despite the fact that training such big networks might be costly. This algorithm was experimentally evaluated in classification, regression, and RL problems recording good results.
- Probabilistic Model-Agnostic Meta-Learning (PLATIPUS) was introduced by Chelsea Finn et al. [12] in 2018. It is a probabilistic algorithm that can sample models for a new task from a model distribution. PLATIPUS also uses a variational lower bound to train a parameter distribution. Moreover, it is simple and provides comparable results to MAML in experimental evaluation. However, it is less efficient in assessing uncertainty because it provides a nearly poor estimation of posterior variance.
- Amortized Bayesian Meta-Learning (ABLM) proposed by Ravi and Beatson [13] in 2018 is a technique that successfully employs hierarchical variational inference to learn a prior distribution. ABLM also provides a high-quality approximation posterior. Compared to MAML and PLATIPUS, the experimental evaluation shows that this model achieves better computation in the uncertainty estimations.

TABLE I
EXPERIMENTAL EVALUATION FIELDS

Algorithm	Classification	Regression	Reinforcement Learning
MAML	x	x	x
MAML++	x	-	-
iMAML	x	-	-
ALPHA-MAML	x	-	-
OOD-MAML	x	-	-
FOMAML	x	-	-
Reptile	x	-	-
HF-MAML	-	-	-
LLAMA	x	x	-
BMAML	x	x	x
PLATIPUS	x	x	-
ABLM	x	-	-

IV. DISCUSSION AND FUTURE DIRECTIONS

All the above algorithms are demonstrated on different model types and are experimentally evaluated in several distinct domains, i.e., classification problems, few-shot regressions, and RL scenarios. Table I indicates that the presented algorithms are mostly designed for classification problems. Some of them are also tested in regression problems and only the first MAML algorithm and its Bayesian format (BMAML) are used in RL tasks.

In contrast to the prior algorithms, other MAML subvariants have also been proposed focused only on reinforcement, deep reinforcement, and meta-reinforcement learning. The Taming MAML [14] using the automatic differentiation framework, and the ES-MAML [15], which is based on Evolution Strategies, are two typical examples of this category.

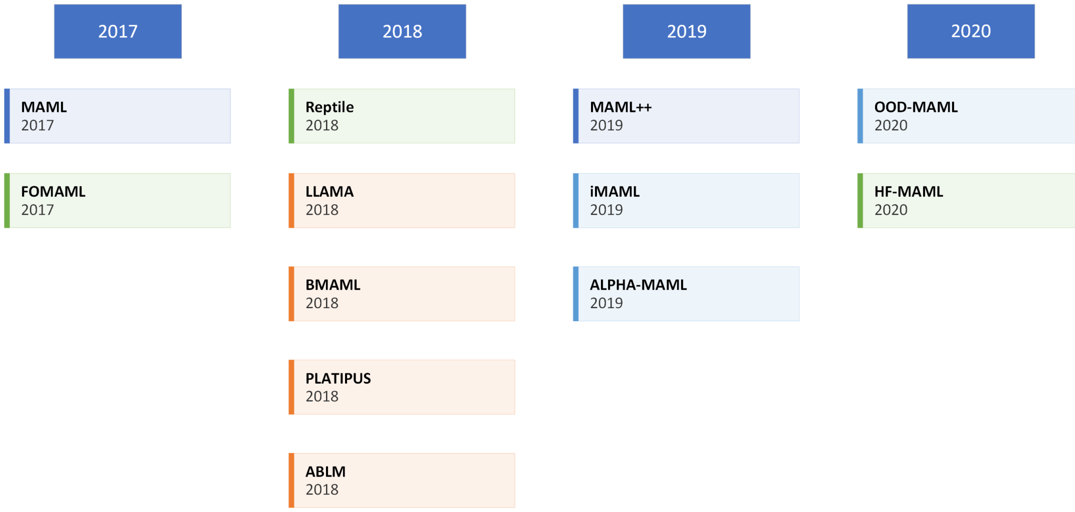

Fig. 2. MAML algorithms in timescale. Different colors are used for each category (Grey for the first MAML algorithm, blue for regular gradient descent MAML algorithms, green for the Hessian-free MAML algorithms, and orange for Bayesian MAML algorithms).

Another observation is that most of the works regarding the MAML algorithm family were proposed between 2017 and 2020 as shown in Fig. 2. At this time, the research interest has moved to new RL methods, which are a widely developing field in the deep-learning field. However, the MAML family consists of simple and widely used algorithms and there is a change that in the next years the researchers will continue to evolve this field.

V. Conclusions

In this work, a brief review of the most popular MAML techniques is presented. Various algorithms are proposed as an extension to the first MAML algorithm to overcome the weaknesses of the main proposal. The main characteristics that researchers try to decode are training stability, memory efficiency, flexibility, generalization, computational complexity, and robustness. Most of these algorithms are experimentally evaluated in few-shot classification problems and achieved promising results. Due to the simplicity and adaptability of the MAML algorithm family, they established a field with further potential development. In future work, we will extend our research to RL techniques. We are planning to make an extensive review as well as an experimental evaluation of the algorithms that are used in this field.

References

[1] T. Hospedales, A. Antoniou, P. Micaelli and A. Storkey, "Meta-Learning in Neural Networks: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5149-5169, 1 Sept. 2022, doi: 10.1109/TPAMI.2021.3079209.

[2] R. Vilalta, C. Giraud-Carrier, and P. Brazdil, "Meta-Learning - Concepts and Techniques", in Data Mining and Knowledge Discovery Handbook, Springer, Boston, MA, pp. 717–731., 2010, doi: 10.1007/978-0-387-09823-4-36.

[3] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in Proc. 34th International Conference on Machine Learning - Volume 70, Sydney, Australia, Aug. 2017, pp. 1126–1135, doi: 10.48550/arXiv.1703.03400.

[4] A. Antoniou, H. Edwards, and A. Storkey, "How to train your MAML," in International Conference on Learning Representations, New Orleans, LA, USA, May 2019, doi: 10.48550/arXiv.1810.09502.

[5] A. Rajeswaran, C. Finn, S. M. Kakade, and S. Levine, "Meta-learning with implicit gradients," Advances in neural information processing systems, vol. 32, Vancouver, Canada, 2019.

[6] H. S. Behl, A. G. Baydin, and P. H. Torr, "Alpha maml: Adaptive model-agnostic meta-learning," in 6th ICML Workshop on Automated Machine Learning, California, USA, 2019, doi: 10.48550/arXiv.1905.07435.

[7] T. Jeong and H. Kim, "Ood-maml: Meta-learning for few-shot out-of-distribution detection and classification," Advances in Neural Information Processing Systems, vol. 33, pp. 3907–3916, Dec. 2020.

[8] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," 2018, arXiv:1803.02999.

[9] A. Fallah, A. Mokhtari, and A. Ozdaglar, "On the convergence theory of gradient-based model-agnostic meta-learning algorithms," in International Conference on Artificial Intelligence and Statistics, pp. 1082–1092, 2020.

[10] E. Grant, C. Finn, S. Levine, T. Darrell, and T. Griffiths, "Recasting gradient-based meta-learning as hierarchical bayes," in International Conference on Learning Representations, Vancouver, Canada, 2018, doi: 10.48550/arXiv.1801.08930.

[11] J. Yoon, T. Kim, O. Dia, S. Kim, Y. Bengio, and S. Ahn, "Bayesian model-agnostic meta-learning," in Advances in Neural Information Processing Systems, vol. 31, 2018, doi: 10.48550/arXiv.1806.03836.

[12] C. Finn, K. Xu, and S. Levine, "Probabilistic model-agnostic meta-learning," Advances in Neural Information Processing Systems, vol. 31, 2018, doi: 10.48550/arXiv.1806.02817.

[13] S. Ravi and A. Beatson, "Amortized bayesian meta-learning," in International Conference on Learning Representations, New Orleans, LA, USA, May, 2018.

[14] H. Liu, R. Socher, and C. Xiong, "Taming maml: Efficient unbiased meta-reinforcement learning," in International conference on machine learning, Long Beach, CA, USA, June, 2019.

[15] X. Song, W. Gao, Y. Yang, K. Choromanski, A. Pacchiano, and Y. Tang, "Es-maml: Simple hessian-free meta learning," in International Conference on Learning Representations, Apr. 2020, doi:10.48550/arXiv.1910.01215.