Understanding deep meta-learning

Huisman, M.

Summary

Deep neural networks have demonstrated impressive performance in various domains, often achieving human-level or super-human capabilities. However, their success heavily relies on the availability of extensive training data, which becomes problematic in domains where data collection is challenging, costly, or privacy-sensitive. Enhancing the data efficiency of deep neural networks and overcoming these limitations is of utmost importance. By doing so, we can unlock the full potential of these networks, enabling them to learn and adapt effectively even when presented with limited data. This improvement also paves the way for their deployment in resource-constrained environments, promoting the democratization of deep learning by making it more accessible and applicable across various domains.

One of the potential causes of their inefficient learning ability is the fact that deep neural networks are often trained from scratch in an end-to-end way. Deep meta-learning has emerged as a rapidly growing field to improve the learning efficiency of deep neural networks by endowing them with the ability to reuse prior experiences and knowledge. This dissertation focuses specifically on the application of deep meta-learning in few-shot learning scenarios, where networks need to rapidly adapt to new tasks with only a limited number of examples.

Despite recent progress in few-shot learning, the underlying principles that drive the success of meta-learning algorithms are still poorly understood, impeding algorithm development and design choices. In this dissertation, we take steps to bridge this knowledge gap by gaining a comprehensive understanding of the fundamental principles of popular deep meta-learning algorithms, enabling more informed algorithm development, and establishing robust theoretical foundations. Additionally, this work explores the integration of theoretical principles into practical algorithm design to enhance the capabilities of deep meta-learning approaches. By addressing these research gaps, this dissertation aims to advance the field, paving the way for more effective and principled meta-learning techniques that offer broader applicability and superior performance. Below is a brief outline of the dissertation.

Chapter 2 serves as an extensive introduction and overview, providing readers with a solid theoretical foundation for understanding deep meta-learning algorithms. We delve into key methods and categorize them into three main categories: i) metric-based techniques, ii) model-based techniques, and iii) optimization-based techniques. By exploring these approaches, we aim to provide a holistic understanding of the diverse methodologies employed in deep meta-learning. Furthermore, we identify and discuss the primary open challenges in the field. These challenges include the need for performance evaluations on heterogeneous benchmarks to ensure the robustness and generalizability of meta-learning algorithms.

In Chapter 3, we investigate an empirically observed performance gap between two popular and highly related deep meta-learning algorithms: the meta-learner LSTM and MAML. We found

this performance gap surprising based on our work in Chapter 2 as the meta-learner LSTM is more expressive than MAML and could, in theory, emulate the behavior of MAML. To gain a deeper understanding of this performance gap, we introduce a novel algorithm called TURTLE. The design and analysis of TURTLE reveal that the notable performance gap can be attributed to the influence of second-order gradients. We find that second-order gradients can also significantly increase the accuracy of the meta-learner LSTM with slight modifications of the inputs provided to the LSTM.

A related method to deep meta-learning is the transfer learning method commonly known as pre-training and fine-tuning. In Chapter 4, we investigate the observed performance differences between finetuning, MAML, and another meta-learning technique called Reptile. We present evidence indicating that MAML and Reptile exhibit a tendency to specialize in rapidly adapting to low-data regimes characterized by similar data distributions as the ones used during training. Our findings highlight the importance of both the output layer and the presence of noisy training conditions induced by data scarcity in the few-shot learning setting. These factors contribute significantly to enabling the specialization observed in MAML and Reptile. Additionally, we demonstrate that the pre-trained features obtained through the finetuning baseline exhibit greater diversity and discriminative power compared to those learned by MAML and Reptile. This lack of diversity and distribution specialization in MAML and Reptile may hinder their ability to generalize effectively to target tasks that differ significantly from the observed training tasks. In contrast, finetuning can leverage the diverse set of learned features to adapt more successfully to such distant target tasks.

In Chapter 5, we revisit a classical LSTM approach to deep meta-learning, where the idea is to feed a training dataset into an LSTM and to condition the predictions of query inputs on the resulting hidden state. This approach is known to be maximally expressive, that is, the LSTM could learn to implement any learning algorithm. Despite the promising results of this approach on small problems and on reinforcement learning problems, the approach has received little attention in the supervised few-shot learning setting. We show that LSTM outperforms the popular meta-learning technique MAML on a simple few-shot sine wave regression benchmark, but that LSTM, expectedly, falls short on more complex few-shot image classification benchmarks. We identify two potential factors contributing to the observed limitations and propose a novel method called Outer Product LSTM (OP-LSTM) to address these issues effectively. OP-LSTM surpasses the performance of plain LSTM and exhibits substantial performance gains. While these results alone do not set a new state-of-the-art, the advances of OP-LSTM are orthogonal to other advances in the field of meta-learning, yielding new insights in how LSTM works in image classification, allowing for a whole range of new research directions.

In Chapter 6, we investigate whether the integration of the fact that more expressive models are more likely to overfit can improve the few-shot learning performance by meta-learning which parameters to adjust. To investigate this, we propose *Subspace Adaptation Prior* (SAP), a novel gradient-based meta-learning algorithm that jointly learns good initialization parameters (prior knowledge) and layer-wise *parameter subspaces* in the form of operation subsets that should be adaptable. In this way, SAP can learn which operation subsets to adjust with gradient descent based on the underlying task distribution, simultaneously decreasing the risk of overfitting when learning new tasks. We demonstrate that this ability is helpful as SAP yields superior or competitive performance in few-shot image classification settings (gains between 0.1% and 3.9% in accuracy). Analysis of the learned subspaces demonstrates that low-dimensional operations often yield high activation strengths, indicating that they may be important for achieving good few-shot learning performance.

As such, in this dissertation, we have analyzed and performed empirical validation of various meta-learning systems, including MAML, Reptile, finetuning and various LSTM-based approaches. Additionally, we have explored the integration of theoretical principles for practical algorithm development. In short, we have made a small step toward understanding deep meta-learning algorithms, paving the way for more robust and principled meta-learning techniques with broader applicability and superior performance.