

The art of Learning to Learn by using the Power of Meta-Learning in AI

Author, Abhishek Mishra

A Data Science Foundation White Paper

March 2020

www.datascience.foundation

Copyright 2016 - 2017 Data Science Foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3641 Email: admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670

Machine learning is a fast-paced area and many types of research are ongoing. One of the fastest growths can be seen in the area of Meta-learning.

As it is becoming more popular and more meta-learning techniques are being developed, it is important to understand this area of data science. Meta-learning caught my attention somewhere back in 2018 and I have seen the techniques being improved, more areas being covered, and many new experiments being implemented which have shown successful results with Meta-learning.

Let's begin to understand the subject from its origin to present usage and implementation. Here are some of the key points I will cover in this paper.

1. What is Meta-Learning?
2. Where did it originate?
3. What are the various approaches?
4. What is its advantage?
5. A Sample approach explanation

Now that you are aware of the above, let's start this with a famous saying:

"YOU ARE ONLY CONFINED BY THE WALLS YOU BUILD YOURSELF"

So, let's break the walls and get started.

Meta-Learning is originally a concept of cognitive psychology

John Biggs (1985) used the concept of meta-learning to describe the state of "being conscious of one's learning and taking control of it." In contrast to subject knowledge, you may characterize meta-learning as an awareness and understanding of the process of learning itself. It is like thinking about thinking.

Donald Maudsley coined the term meta-learning to describe a mechanism by which people begin to influence what they learn, becoming "increasingly in charge of the patterns of perception, inquiry, learning, and development that they have internalized."

The basic concept of "learning how to learn" is extended to AI systems for the machine learning component of meta-learning. Ideally, in the Machine learning world, it is defined as a task where the machine learns how to do various tasks of different complexity while the overarching principle that when one task is learnt, the knowledge of how that was learnt, is then applied to other tasks.

This can be done only when Machine learning models are trained to do smaller subtasks, this kind of training takes a long time and the way many AI agents are designed it is not easy to transfer knowledge learned during one task to another task. Thereby giving a huge opportunity for meta-learning to bridge this knowledge transfer gap and this is where meta-learning plays a vital role. Meta-learning models and techniques help the AI not only to learn fast but also how to generalize learning methods and gain new skills faster.

We are not focused on label acquisition in the world of meta-learning; rather we are trying to create a system that learns quickly from a limited amount of training data. Unless we go by the concept of social psychology, meta-learning is the state of being conscious of one's learning and of taking control. When applied to the machine learning theory, similar ideas state that a meta-learning algorithm uses prior experience to modify certain aspects of an algorithm, so the changed algorithm is better than the original algorithm. In simple terms, Meta-learning is how the algorithm learns how to think and learn.

What is meta-learning trying to do?

We human beings learn new concepts, ideas, and abilities or skills quickly and effectively. However, machine learning models often require training on the large dataset for the same. Children who have seen snakes and dogs just a few times can differentiate them quickly. We can apply the same principles and design machine learning models with similar properties that will enable the machine to learn a skill quickly and with few training examples. This is what meta-learning is trying to do. We expect that a good meta-learning model will adapt or generalize well to new tasks and new environments that were never experienced during the training period. Usually, these mini learning sessions take place in the testing phase of machine learning models with minimum exposure to new task parameters. That is the prime reason why meta-learning is also known as learning to learn.

Here are a few examples of meta-learning tasks:

1. A classifier that is trained on a handful of dog images can tell whether a given image contains a dog or not.
2. Imagine you playing a chess game online with a computer and the bots easily beat you, so you switch to cards, and that AI bot is still able to play quite well in that game. A bot or precisely a game bot can excel in a new game by using its knowledge of past games and strategy and with its ability to learn how to learn.

In the above examples, Training of the parameters in the model has been performed in such a way that a small number of gradient changes contribute to faster learning on a new task.

I hope this helps to get a hang of it.

The model is going to be hungry for data and forced to learn less about data. Two strategies are considered widely in this scenario:

1. Reuse of functionality
2. Learning faster

Reuse of the functionality includes meta initialization of high-quality features already in place and learning faster is the utilization of massive and efficient changes in the representations.

In learning faster all the major parameter changes take place during adaptation of each new task, as a result, a conditional weight is assigned from meta initialization, while in reuse of functionality the meta initialization already has the most useful features that can be reused for a new task.

To understand whether meta-learning models benefit from either of the two, a few researchers from Google, MIT and Cornell University have collaborated to evaluate the results and suggest that reuse of functionality is the dominating factor of improving the effectiveness of meta-learning algorithms.

Read more in the paper presented by the authors:

<https://arxiv.org/pdf/1909.09157v1.pdf>

Meta-learning implementations are not limited to only semi-supervised tasks but can be used in tasks such as the selection of objects, estimation of densities and reinforcement of learning tasks.

Three main steps to create a meta-learning model are:

1. Extract relevant and useful knowledge and experience from an ml model
2. Include learning submodels
3. Allow for inductive dynamic bias

Now, let us see what the advantages of Meta Learning are

1. Less data required to train: These methods help to build more general frameworks that can transfer information from one context to another. This reduces the amount of data you need in the new context to solve the problems.
2. Speed: Meta-learning ways and method help in producing custom made models which perform better at a high speed.
3. Scalable: Meta-Learning models help to increase the scalability of an AI application by automating the process and improving algorithms.
4. Agile and adaptable to environmental changes like Reinforcement learning.

Meta-learning algorithms are already used in different applications, some of them being:

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3641 Email: admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670

1. As discussed above, image classification tasks
2. Detecting fraudulent transactions
3. Auto-detection of placeholders in images.

Great, now we know about a few basic uses of meta-learning algorithms, now let us discuss some specific types of meta-learning algorithms:

Few-Shots Meta-Learning

A few-shots meta-learning approach is one where a deep neural networks are designed in such a way that they can generalize to unknown data sets from the training datasets. A few-shot classification example is similar to a standard classification task, but where the data samples are whole datasets. The model is trained on many different learning datasets and then on the assembly of training tasks and unknown data it is optimized for peak performance, this leads to a split of single training samples into multiple classes.

The idea is that the individual training samples are lightweight in few-shot learning and that the network can learn to identify objects after seeing only a few images and it is like how a child learns to identify objects after only seeing a few pictures. This methodology was used to build techniques such as one-shot generative models and memory-enhanced neural networks.

Few-Shot learning aims to identify new data after seeing only a few examples of training. Few-shot learning offers a potential solution by learning to learn across data from many previous tasks.

Optimizer Meta-Learning

Meta-learning is also used to improve the efficiency of a neural network already in use. Meta-learning methods for optimizers typically work by adjusting the hyperparameters of a single neural network to improve the performance of the base neural network. The consequence is that the target network should get better at carrying out the task on which it is being trained. Models which focus on improving gradient descent techniques are a great example of optimizer meta-learning

Metric Meta-Learning

Metric meta-learning aims to determine a metric space where particularly successful learning is found. Metric-based meta-learning is the use of neural networks to assess whether a metric is being used successfully and whether the network or networks reach the goal metric. It is similar to few-shot learning, as only a single example is used to train the network and have the metric space learned. Throughout a perse domain, the same metric is used and if the networks perge from the metric they are considered to

fail.

Recurrent Meta-Learning

Recurrent meta-learning paradigm is the application of meta-learning methods to Recurrent Neural Networks (RNN) and related networks of Long Short-Term Memories (LSTM). The strategy is to train the RNN / LSTM model to sequentially learn a data set and then use this learned model as a basis for another learner. The meta-learner takes into account the particular algorithm of optimization used to train the initial model. The meta-learner's inherited parameterization helps it to initialize and converge easily but can still upgrade for the new task.

Ensemble Methods

These methods combine 3 basic types; Boosting, Bagging and Stacked generalization

The way Meta-Learning works depends upon the model and nature of the task in hand, ideally it's like a passage of first network parameters into the second network. The meta-learning model is usually trained after several training measures have been performed on the base model. The forward training pass for the optimization model is carried out after the forward, backward, and optimization steps which train the base model. After the meta-loss is determined for each meta-parameter the gradients are calculated which are changed in the optimizer. One way to calculate the meta-loss is to finish the initial model's forward training pass and then add the losses which have already been measured.

In the case of a set of algorithms, each instance gathers features or meta-functions, a single value characteristic of the task, information like:

Simple (number of rows, columns) of the task-representing dataset, statistics (average features: mean, standard deviation, skewness), theoretical information such as entropy, just to name a few; the concern is to learn from experience through multiple tasks.

Any meta-learning algorithm and its variants essentially tend to be fully self-referential. That means it can inspect and enhance each part of its code automatically.

Let us begin to understand this with an example from Walmart labs on Fraudulent transaction detections

They use a meta learner to detect fraudulent transactions based on the tasks which the low-level models are trained on. Here is what they have to say:

"Meta-learning could be used to resolve the use cases we mentioned above when there are only 10 to 100

training examples. The top-level model tunes the bottom-level models (each model could be made from a different task) to extract knowledge from them. Then, it predicts the results used to create new training examples. If its prediction is correct, then the teacher would reward it; otherwise, it is penalized. In this situation, the teacher refers to the optimizer to penalize the weights of the top-level model (student).

So, meta-learning takes knowledge from previous tasks to create a solution for the current task. Thus, it aids in optimizing the low-level AI model's architecture, hyperparameters, and dataset tuning."

Second case, 'Detecting placeholder in images', here is their approach:

"We may have multiple models for doing n number of tasks based on classification, regression, and Reinforcement Learning tasks based on Q-learning, Double Q-learning, etc. Now, given an image to check if it is a placeholder or not, the meta-learner being the top-level AI model will try to infer the predicted value from the bottom-level AI models, like models built on image classification, image-based QA to Reinforcement Learning tasks based on Q-learning, Double Q-learning, etc. Let's say our meta-learner gives a probability number of whether the image is a placeholder or not.

Now, the meta-learner acts like a student and rewards itself if the predicted result is correct. If it's not correct, it is penalized by the teacher (the actual value). Hence, with the help of just one example, your meta-learner tries to learn and gradually provides the correct results. This is an example of few-shot learning that does not require labeled data."

The above is an extract from <https://medium.com/walmartlabs/an-introduction-to-meta-learning-ced7072b80e7>

MAML (Model Agnostic Meta-Learning)

Ok, now we have a good idea of the ecosystem, so let's look at MAML. It is a breakthrough in the meta-learning research area and has created a of interest.

MAML's basic idea is to find better initial parameters so that the model can learn quickly on new tasks with a lower gradient step.

Let's say we do a classification function using a neural network. How are we training the network? We must initialize random weights and train the network by minimizing the loss. But how do we do loss minimization?

Ideally, we use gradient descent for coming up to an optimal weight that will give us minimal loss, by taking multiple gradient steps we reach the convergence. In MAML, we try to find optimal weights by

learnings from the distribution of similar tasks. So, for a new task, then we don't have to start with randomly initialized weights, we can start with optimal weights that will take lower gradient steps to reach convergence and with no more data points being needed for training.

In MAML, we minimize the loss using gradient descent by sampling a batch of the task and getting optimal parameters thereby when we sample another batch, we update our randomly initialized model parameters by calculating the gradient concerning optimal parameters in a new set of tasks.

Meta-Learning is a vast topic and it covers many aspects of various related domains. While meta-learning methods are currently computationally expensive, they are an exciting frontier for AI Research and can be a significant step forward in our quest for Artificial Intelligence, because computers will not only be able to make accurate classifications and estimates but will also be able to improve their parameters (and hyperparameters) to better perform multiple tasks.

Below are a few learning resources for you to review and gain better knowledge about meta-learning.

Learning resources:

<http://nn.cs.utexas.edu/?neuroevolution-tutorial-ijcnn2013>

<https://blog.coast.ai/lets-evolve-a-neural-network-with-a-genetic-algorithm-code-included-8809bece164>

<https://towardsdatascience.com/paper-repro-deep-neuroevolution-756871e00a66>

References:

Giraud-Carrier, Brazdil, P., C., Soares, C. and Vilalta, R. (2009). Met- learning: Applications to Data Mining, Springer-Verlag.

Rice, John R., (1975). The Algorithm Selection Problem, Computer Science Technical Reports. Paper 99.

https://en.wikipedia.org/wiki/Meta_learning

[https://en.wikipedia.org/wiki/Meta_learning_\(computer_science\)](https://en.wikipedia.org/wiki/Meta_learning_(computer_science))

<https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html>

Data Science Foundation

<https://arxiv.org/pdf/1909.09157v1.pdf>

<https://www.borealisai.com/en/blog/tutorial-3-few-shot-learning-and-meta-learning-ii/>

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: admin@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3641 Email: admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670