Maria Doliashvili

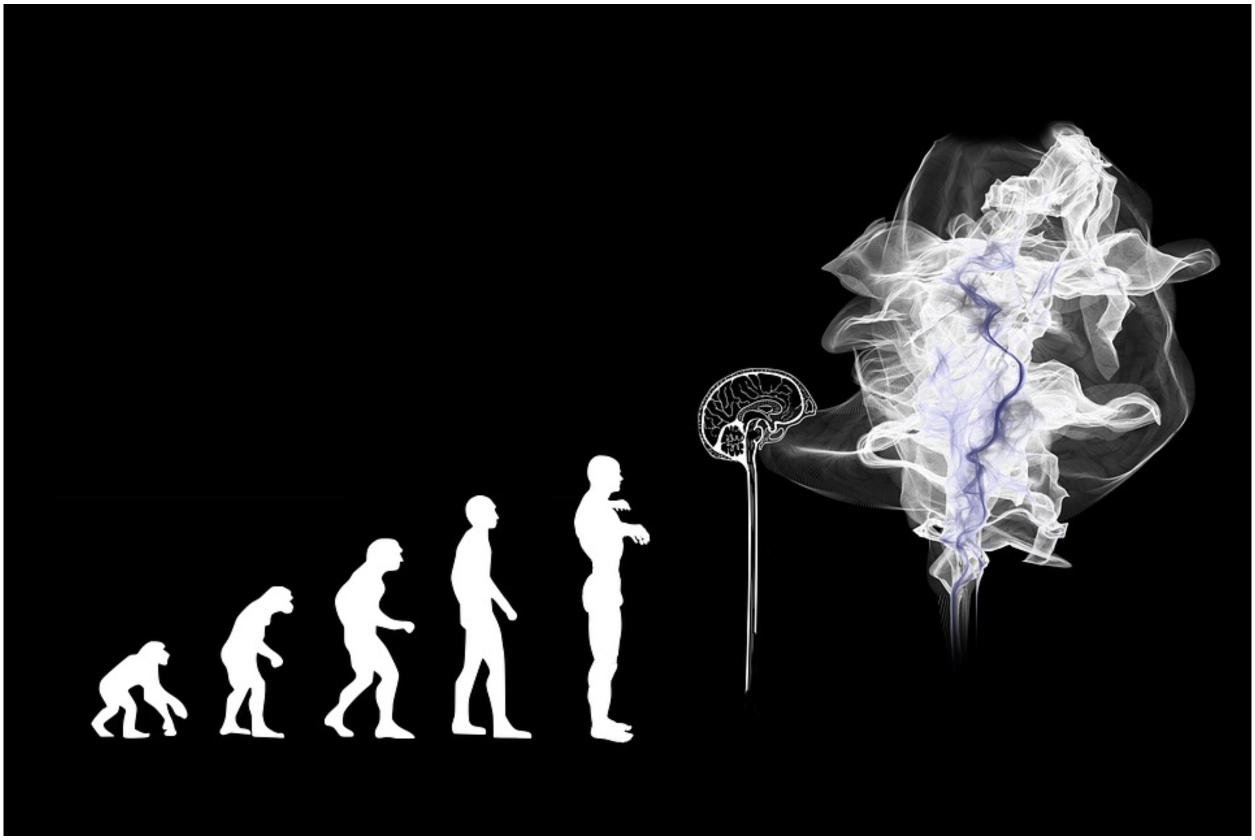
Learning to Learn - Artificial Evolution

Meta-Learning

Abstract

Learning to learn by solving many tasks is called meta-learning. The idea behind meta-learning for neural networks is to train a system on many tasks with the goal that it can solve new tasks and adapt to new environments quickly; by observing how different machine learning approaches perform on a wide range of learning tasks. Current artificial intelligence systems excel at constructing a single approach for a singular task, however fail to generalize and perform adequately when the task or the approach is being altered. There are several promising meta-learning approaches being developed and several success stories that will be surveyed in this literature review.

Keywords: Meta-learning, AutoML, Meta-knowledge, Hyperparameters, Few-Shot Learning



Introduction

"... a system that improves or discovers a learning algorithm"

(Hochreiter, Younger, & Conwell, 2001)

Human beings acquire knowledge about how to perform each task and re-use some of the knowledge in the future for performing new tasks or adapting to new environments. As the set of previously learned tasks increases, humans learn new skills more and more effectively. For example, learning to distinguish each animal from each other (like penguins from red pandas) only takes up to a few examples for training, while interacting with strangers can be an overwhelming experience for an infant, it becomes easily manageable for an adult. As for artificial intelligence systems nowadays, they almost always learn how to perform a task starting from scratch each time for every new task and environment. None of the previous approaches that worked well are being reused. Trainings for building a good machine learning model usually take a large amount of data. For example machine translation models cannot hold a dialogue or any kind of conversation and DeepMind's AlphaGo while beating a master player Li Sidol in board game Go, cannot help Boston Dynamics robots in parkour.

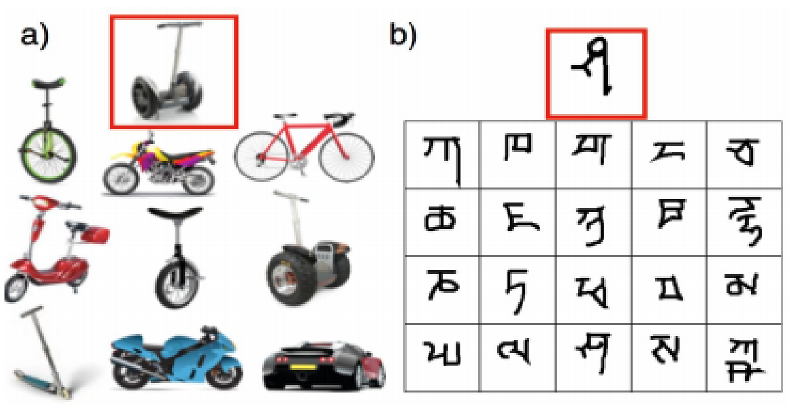
When we (humans) build machine learning models, we usually use our previous experiences of building them and remember and reuse the approaches/models that worked well on similar tasks. We use settings that previously worked while we hand tune the parameters, etc. Basically we are training ourselves on how to build a machine learning model. The main task of meta-learning can also be phrased as: Is it possible to create a system that will be able to automate the process of creating optimal models? And can these systems be more efficient than human beings? How can we let AI models to be as adaptive to new environments and re-use the previous learning knowledge the same way humans do?

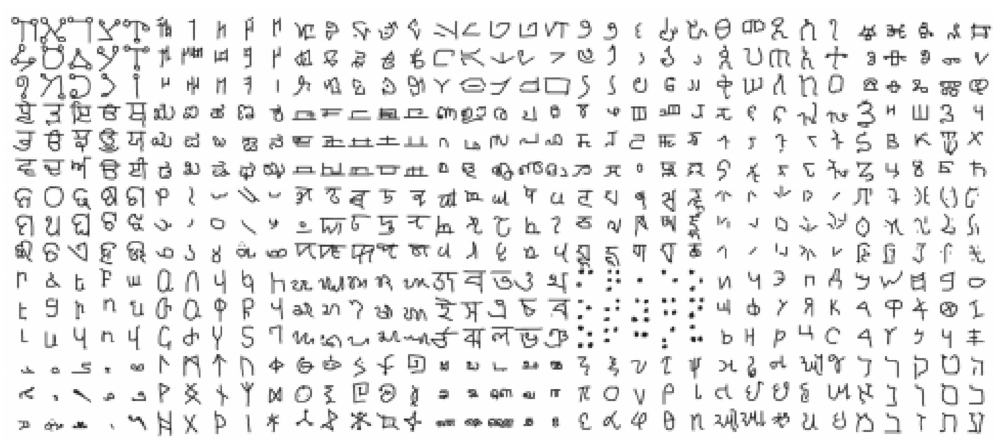
There are several techniques attempting to address the above questions. For creation of intelligent agents that are able to perform several tasks and acquiring several skills, it would be inefficient if we allocate the time and resources needed for learning each skill/task from scratch. Instead aiming to learning to learn the agents should not consider each task separately from each other, the tasks should be the training cases simultaneously for the model. The adaptation process should be done reusing the previous experience and it should make it possible to spend only a little time on training and with limited information about the new task or the environment that it did not encounter during the training time. The adapted model should be able to complete the new tasks. Basically, we can also look at the current approaches to meta-learning as a type of deep learning where instead of training cases we have training tasks and instead of test cases we have test tasks. Therefore meta-learning should enable us to create models that will be able to solve more complicated and interdisciplinary tasks in the future by providing more adaptive models and by doing so allowing us to be one step closer to artificial general intelligence.

The first meta-learning approaches are from the 1980s, though there are several remarkable papers from the recent years that made the topic more popular. Recent work includes techniques using hyperparameter and neural architecture search, few shot image recognition and reinforcement meta-learning, etc. (Bengio, Bengio, & Cloutier, 1990; Schmidhuber, 1987).

Few-Shot Learning

Lake et al. published a paper in 2015, challenging machine learning models to learn a task given only few instances of the task (Lake, Salakhutdinov, & Tenenbaum, 2015).





Lake et al, 2013, 2015

Lake suggested that humans can identify novel two-wheel vehicles from a single picture and draw a character in a new alphabet, after seeing a few examples, while machine learning models cannot generalize any concept from few examples. Lake et al. also introduced Omniglot - a handwritten version of MNIST dataset, with 1623 (20 examples) character classes (Lake, Salakhutdinov, & Tenenbaum, 2013). It was shown in 2016 that it is possible for deep learning models to learn to learn from the few examples on Omniglot dataset (Lake, Ullman, Tenenbaum, & Gershman, 2017). Few shot object recognition became a hot topic after the initial success with meta-learning. That is essentially selecting hyper-parameters and parameters that can adapt to a new task without overfitting given the few shots, and resulted in one of the top benchmarks for few shot learning.

Some of the meta-learning methods for few-shot learning that should be mentioned are Lee et a.l using embeddings and SVM; edge-labeling graph neural network for few-shot learning by Kim et al. which represents each image as a node in a graph and there are features per edge based on how similar the nodes are to each other; task agnostic meta-learning for few-Shot learning by Jamal et al.; meta-transfer learning for few-shot learning by Sun et al., etc (Jamal & Qi, 2019; Kim, Kim, Kim, & Yoo, 2019; Lee, Maji, Ravichandran, & Soatto, 2019; Sun, Liu, Chua, & Schiele, 2019).

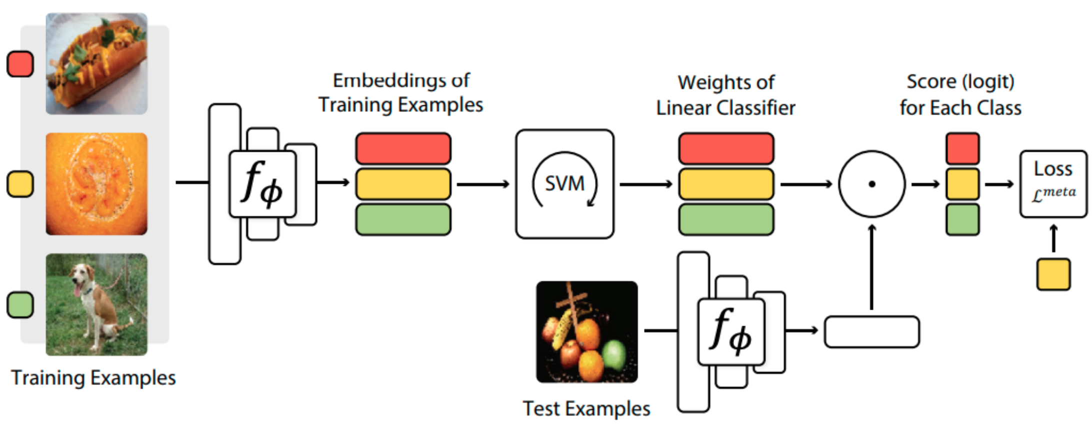


Figure 1. **Overview of our approach.** Schematic illustration of our method MetaOptNet on an 1-shot 3-way classification task. The meta-training objective is to learn the parameters ϕ of a feature embedding model f_ϕ that generalizes well across tasks when used with regularized linear classifiers (*e.g.*, SVMs). A task is a tuple of a few-shot training set and a test set (see Section 3 for details).

Neural Architecture Search

Neural Architecture Search (NAS) is related to both AutoML and hyperparameter optimization (Elsken, Metzen, & Hutter, 2018; Zoph & Le, 2016). It is a way of optimizing designing a neural network architecture. NAS is shown to be able to design networks that are as good or even outperform hand-designed architectures. We can formulate the task as follows: Find a new network architecture with only a few bits to specify. Learn it on a small dataset, test it on a large one.