Meta Learning

Gianluca Guglielmo

s238906@outlook.com

Introduction

Meta-Learning refers to the process of learning to learn. It doesn't have a standard interpretation yet, since different techniques can be considered as fitting: fine-tuning, optimization-based and metric-based meta-learning are some examples. What they have in common is the exploitation of meta-knowledge gained from previous experiences. Humans are naturally capable of doing so [1]: our brains can easily generalize new tasks and use previously learned skills to create a fast training-test experience, rather than starting from scratch. For instance, it's safe to assume that each case in which we need to move an object from point A to point B is fairly new: the environment changes, along with the characteristics of the object and our general condition. This doesn't require us to learn to move an object altogether: we simply quickly assess the situation (*Few-Shot Learning*) and then accomplish the goal.

Generalization is not automatic in ANNs, but it can be reached with some careful planning. Given a *Dataset D* and a *Loss function* $\mathcal{L}$, a *Task* $\mathcal{T}$ is a Machine Learning procedure that outputs a model *function* f_Θ. More formally, a *Task* $\mathcal{T}$ is defined as:

$$\mathcal{T} \triangleq \{p_i(x), p_j(y|x), \mathcal{L}\}$$

The Meta-Learning problem can be defined as: given data about from $\mathcal{T}_1 \dots \mathcal{T}_n$, quickly solve new task $\mathcal{T}_{test}$.

Extracting prior distributions from previously learned tasks and using it to solve new problems is at the heart of the probabilistic view of meta-learning. Moreover, these new problems can be solved using minimal training by exploiting an internal representation that is broadly suitable for many cases. Tasks must share some structure in order for meta-learning approaches to be useful, which formally translates in: meta-training tasks and meta-test task must be drawn i.i.d. from same task distribution $p(\mathcal{T})$. If this hypothesis is not respected then standard learning approaches are preferred.

A Siamese Neural Network can recognize similarity between inputs [2]. It is a binary classifier that takes two objects as input and outputs 1 if they refer to the same entity, 0 if they do not. This architecture is used in one-shot learning to avoid building a new model when new entities are introduced. It is commonly implemented for face recognition algorithms.

Abstract

In this work I will review the latest breakthroughs in literature regarding Meta-Learning to build a small conceptual map of the different techniques proposed.

Subsequently, I will try to implement a Siamese Neural Network in PyTorch for one-shot learning on the Omniglot's dataset [3]. It is a large dataset of hand-written characters, based upon 50 alphabets from different countries, that contains both images and strokes data.

References

[1] Chelsea Finn. *Learning to Learn with Gradients.* PhD thesis, University of California, 2015.

[2] Gregory Koch. Siamese neural networks for one-shot image recognition. *ICML Workshop*, 2015.

[3] Brenden Lake. Omniglot's dataset. `https://github.com/brendenlake/omniglot`, 2015.