AutoML: Meta-Learning

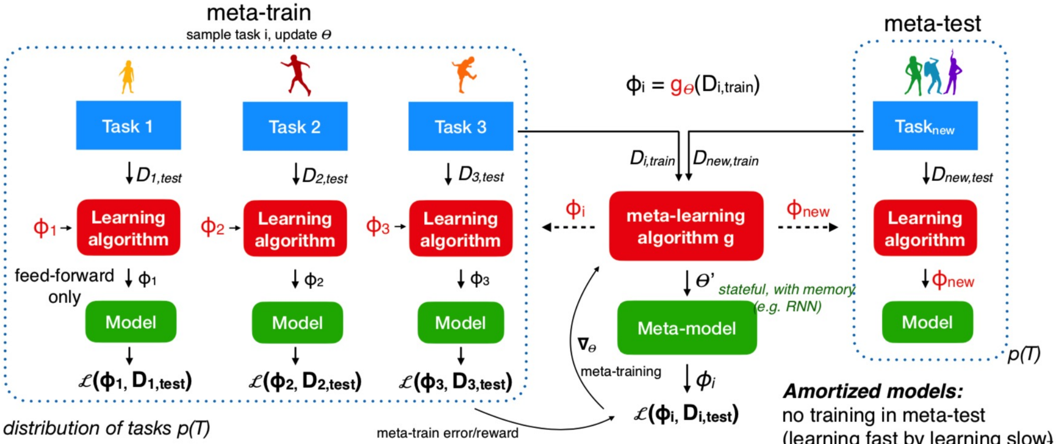
Black-box meta-learning

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

Black-box meta-learning

black box meta-model g_θ predicts ϕ given D_{train} ($theta$ is hidden)

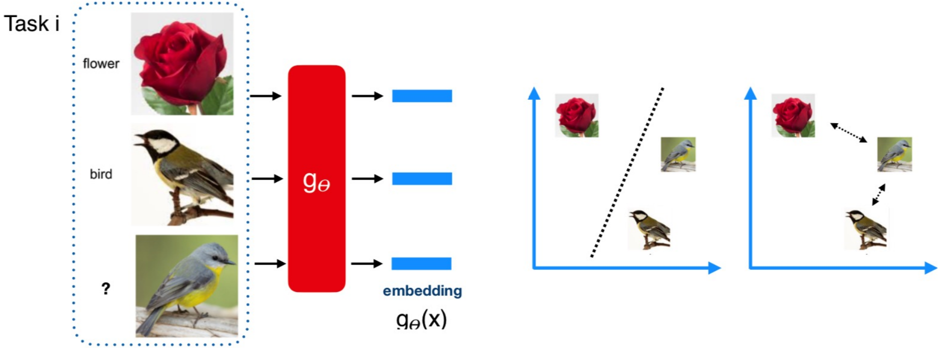
hypernetwork where input embedding learned across tasks



Amortized models:
no training in meta-test
(learning fast by learning slow)

Metric learning

Learn an embedding network $theta$ that transforms data D_{train}, D_{test} across all tasks to a representation that allows easy similarity comparison



Can be seen as a simple black box meta-model

- Often non-parametric (independent of ϕ)

Prototypical networks

- Use an embedding function f_θ to encode each data point
- Define a prototype v_c for every class c based on the examples of that class $D_{train}^{(y)}$

$$\mathbf{v}_c = \frac{1}{|D_{train}^{(y)}|} \sum_{(x_i,y_i)\in D_{train}^{(y)}} f_\theta(x_i)$$

- Class distribution for input x is based on inverse distance between x and prototypes

$$p(y = c|x_{test}) = softmax(-d_\phi(f_\theta(x), v_c))$$

- Distance function can be any differentiable distance
  - E.g. squared Euclidean
- Loss function to learn the embedding:

$$\mathcal{L}(\theta) = -log p_\theta(y = c|x)$$

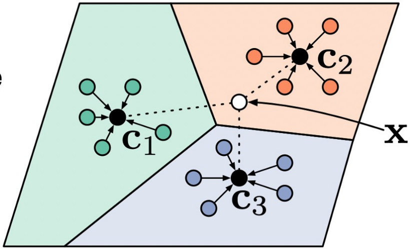
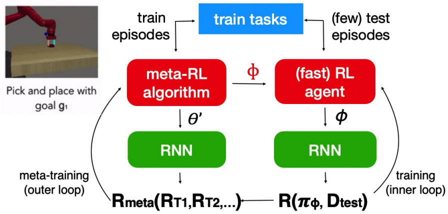


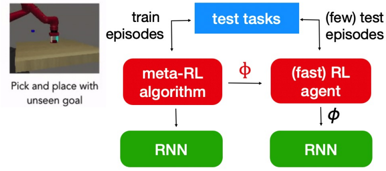
Figure source: Snell et al. 2017

Metric learning

- Quite a few other techniques exist
  - Siamese neural networks
  - Graph Neural Networks
    - Also applicable for semi-supervised and active learning
  - Attentive Recurrent Comparators
    - Compares inputs not as a whole but by parts (e.g. image patches)
  - MetaOptNet
    - Learns embeddings so that linear models can distinguish between classes
- Overall
  - Fast at test time, although pair-wise comparisons limit task size
  - Mostly limited to few-shot supervised tasks
  - Fails when test tasks are more distant: no way to adapt

Black-box model for meta-RL





- RNNs serve as dynamic task embedding storage
- Maximize expected reward in each trial
- Very expressive, perform very well on short tasks
- Longer horizons are an open challenge

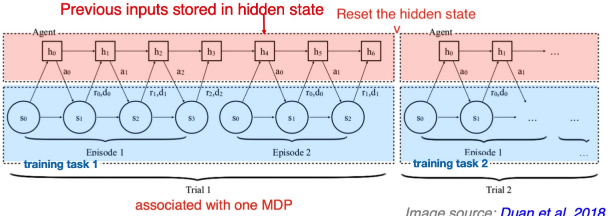
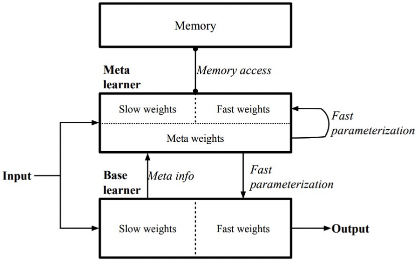


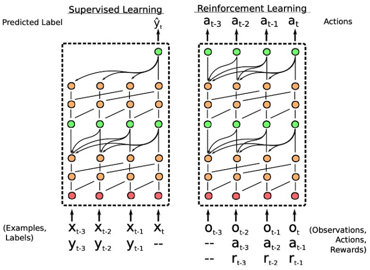
Image source: Duan et al. 2018

Other black-box models

- Memory-augmented NNs
  - Uses neural Turing machines: short term + long term memory
- Meta Networks
  - Meta-learner that returns 'fast weights' for itself and the base network solving the task
- Simple Neural attentive meta- learner (SNAIL)
  - Aims to overcome memory limitations of RNNs with series of 1D convolutions



Meta Networks. *Image source: Munkhdalai et al. 2017*



SNAIL. *Image source: Mishra et al. 2018*