Online Structured Meta-learning

Huaxiu Yao¹, Yingbo Zhou², Mehrdad Mahdavi¹, Zhenhui Li¹, Richard Socher², Caiming Xiong²
¹Pennsylvania State University ²Salesforce Research

Background & Motivation

Structured Prior (e.g. [1]) generalizes better than global prior (e.g., [2])

More realistic scenario — continual prior adaptation (our goal)

Methodology

Block-level Disentangle

Potential New Blocks

Block Search Process:

- Inner
  - $g_{i} = \sum_{j}^{d_{i}} \frac{\exp(a_{ij})}{\sum_{i'}^{d_{i}} \exp(a_{ij'})} M_{i}(w_{0i, t} | g_{i-1, t})$
  - $M_{i}(w_{0i, t}) = w_{0i, t} - o \nabla w_{0i, t} L(w_{0i, t}; D_{i}^{\text{train}})$

- Outer
  - $w_{0i, t} \leftarrow w_{0i, t} - \beta_{i} \nabla w_{0i, t} L(w_{0i, t}; D_{i}^{\text{train}})$
  - $o \leftarrow o - \beta \nabla o L(w_{0i, t}; D_{i}^{\text{train}})$

Finetune: $w_{0i, t}^{\text{FT}} = w_{0i, t} - \beta_{i} \nabla w L(w; D_{i}^{\text{train}} + D_{i}^{\text{test}})$

Experiments

Homogenous: Rainbow MNIST

Data:
- MNIST with shape, color and size change
- 56 tasks in total, 10-way classification

Overall Performance: FTML [3], DPM [4]

Learning Efficiency:

Heterogenous: Meta-dataset

Data:
- Three sub-datasets: Aircraft, Flower, Fungi
- 5-way classification with non-overlap classes, 60 tasks

Overall Performance:

Analysis

➢ Model capacity: increase model capacity to the final size of OSML

Selected ratio of knowledge blocks

References