

CS 330 Project Proposal

Brandon McKinzie & Xiaoying Pang

October 14, 2020

We are interested in building upon techniques for continual meta-learning, wherein the goal is to learn from streams of tasks online/sequentially without forgetting knowledge acquired from the previous tasks. Specifically, our objective is to build a system that meets the following requirements:

1. Learn streams of tasks sequentially [1].
2. Continually learn out-of-distribution (OOD) tasks at deployment/test time [1, 4, 3].
3. Automatically detect task/context switches, such that no explicit task segmentation is needed [3].
4. Learn efficiently without storing data from previous tasks [3, 4, 2].
5. Ability to incorporate changing structure (such as additional class labels, i.e. increasing the “N” for N-way classification).

Existing techniques are able to achieve one or more of these goals, but to the best of our knowledge there is no existing approach that satisfies all points above. OSAKA (or Continual-MAML) tries to learn tasks consisting of out-of-distribution (OOD) data at continual learning (CL) time [3]. However, the fundamental structures of the tasks (i.e. the number of ways/classes of the classification problem) are the same at both training and CL time. Also the OOD datasets they used during test time are still within the same domain as the pretraining data which might be limiting for real-world applications. We therefore propose to extend the definition of OOD tasks to include tasks that 1) have different numbers of ways/classes than what the model is trained on 2) with out-of-domain data much more different from the training data.

In order to extend OSAKA to tasks with different numbers of ways at CL time, we propose to replace the baseline MAML algorithm used in OSAKA with Proto-MAML [5] which is equivalent to first-order MAML and more flexible than MAML in terms of extending to a different number of ways/classes because the task-specific linear layer of the Proto-MAML classifier is initialized from the Prototypical Network equivalent weights. With a set of more diverse tasks, storing and adapting all the accumulated knowledge in a single model like what OSAKA and many other methods do may no longer be sufficient. To make the algorithm learn out-of-domain tasks effectively at CL time, we might want to maintain an ensemble of models. Following the approach of MOLe [4], we’d update the models using expectation maximization (EM). This would accomplish minimum negative transfer between distinct tasks and within each task. For evaluation, we’d track standard loss and accuracy metrics for classification tasks such as on the omniglot dataset. We could also track finer-grain metrics like precision/recall when extending tasks with more classes in order to track the model’s ability to incorporate knowledge on new classes while retaining what it has learned for previous ones. Finally, following OSAKA, we’ll evaluate cumulative accuracy during the entire CL process in order to more accurately measure lifetime performance.

References

- [1] C. Finn et al. *Online Meta-Learning*, 2019.
- [2] J. Harrison et al. *Continuous Meta-Learning without Tasks*, 2019.
- [3] M. Caccia et al. *Online Fast Adaptation and Knowledge Accumulation: a New Approach to Continual Learning*, 2020.
- [4] A. Nagabandi et al. *Deep Online Learning via Meta-Learning: Continual Adaptation for Model-based RL*, 2019.
- [5] E. Triantafillou et al. *Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples*, 2019.
- [6] Kirkpatrick et al. *Overcoming catastrophic forgetting in neural networks*, 2016.