Favorite Paper Summaries

Author: Jaskirat Singh
jaskirat.singh@anu.edu.au

Abstract

While I don't have a strict set of favorite research papers, I believe that my affinity towards a research paper is highly determined by the extent to which I am able to apply the presented ideas from a paper in my own research, or whether I am able to form multiple novel interpretations for the introduced concepts. The following is a list of representative research papers overlapping with my research interests.

Contents

Preprint. Under review.

1 Meta-RL

1.1 Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks (MAML)

Conference. ICML 2017, Finn et al. (2017)

Objectives. For any general collection of tasks, in RL or supervised learning, find an initialization of network parameters θ such that gradient descent updates from that point on a new task lead to quick learning/adaptation.

Merits.

- **Novelty.** The paper proposes to model the *adaptation* step in meta-learning through an optimization procedure. In contrast, prior works take a black-box approach and model the adaptation function using a recurrent neural network.
- **Generalization.** In contrast to black-box approaches, using an optimization based adaptation procedure allows for more flexible adaptation to out-of distribution states that were not encountered during meta-training, thus leading to a better generalization performance.
- **Model-agnostic.** The proposed meta-learning framework can be applied in conjunction with any model design that uses gradient decent for optimization.
- **Multiple interpretations.** In addition to interpreting the proposed approach as an optimization based alternative to performing the inner loop in meta-learning, the intuition for MAML can also be justified using the following mathematical interpretation.

  Given a policy $\pi_\theta(a|s)$ with parameter θ, distribution of tasks $\mathcal{T}_i \sim p(\mathcal{T})$, and task-specific losses $\mathcal{L}_i$, the adaptation and loss computation steps for MAML can be written as follows,

$$\text{Adaptation:} \quad \theta_i' \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}_i(\theta) \tag{1}$$

$$\text{Overall Loss:} \quad \mathcal{L}(\theta) = \sum_i \mathcal{L}_i(\theta_i') \tag{2}$$

  Performing a first-order Taylor series expansion for the overall loss,

$$\mathcal{L}(\theta) = \sum_i \mathcal{L}_i(\theta_i') \tag{3}$$

$$= \sum_i \mathcal{L}_i(\theta - \alpha \nabla_\theta \mathcal{L}_i(\theta)) \tag{4}$$

$$\approx \sum_i \mathcal{L}_i(\theta) - \alpha (\nabla_\theta \mathcal{L}_i(\theta))^2 \tag{5}$$

  Thus, we see that the overall loss function tries to minimize the sum of task-specific losses (which is the usual multi-task approach), and in-addition aims to maximize the sensitivity of the losses to the changes in θ, which is expressed in the second term as the square of gradient of the task-specific loss functions.

2 Visual Navigation

2.1 Learning to Learn How to Learn: Self-Adaptive Visual Navigation using Meta-Learning

Conference. CVPR 2019, Wortsman et al. (2019)

Objectives. The authors use the continuity of the human learning process to argue that an agent must learn from the environment interactions at both train and test times. Thus, they propose to learn a loss function (*interaction loss*) that mimics the gradients of the navigation loss at training time.

Merits.

1. **Novelty.** The paper proposes a continuous learning mechanism, wherein the agent learns from environment interactions during both train and inference times.

2. **Effective use of meta-learning.** The authors propose an effective meta-learning procedure to learn an *interaction loss*, used for self-supervised adaption to unseen test environments.
3. **Comparison with MAML.** The final training objective makes exemplary use of the mathematical interpretation of MAML (discussed above), to minimize both the training-time navigation loss and learn an inference-time interaction loss. To see this, first consider the used training objective,

$$\min_{\theta} \sum_{\tau \in \mathcal{T}_{train}} \mathcal{L}_{nav} \left(\theta - \alpha \nabla_{\theta} \mathcal{L}_{int}(\theta, \mathcal{D}_{\tau}^{int}), \mathcal{D}_{\tau}^{nav} \right) . \tag{6}$$

Taking a first order Taylor expansion, we can decompose the objective as follows,

$$\approx \min_{\theta} \sum_{\tau \in \mathcal{T}_{train}} \mathcal{L}_{nav}(\theta, \mathcal{D}_{\tau}^{nav}) \quad - \quad \alpha \left\langle \nabla_{\theta} \mathcal{L}_{nav}(\theta, \mathcal{D}_{\tau}^{nav}), \nabla_{\theta} \mathcal{L}_{int}(\theta, \mathcal{D}_{\tau}^{int}) \right\rangle \tag{7}$$

Thus, we clearly see that the overall objective aims to minimize the navigation loss on the training tasks (first term), while maximizing the inner product / similarity between gradients for the navigation and *interaction* loss (second term).

3 Generalization in RL

3.1 Quantifying Generalization in Reinforcement Learning

Conference. ICML 2019, Cobbe et al. (2018)

Merits. While not highly mathematical, this paper (along with work by Packer et al. (2018)) provides a great foundation on factors affecting generalization in reinforcement learning.

- **Need for more stochastic RL environments.** The authors show through experiments on procedurally generated game environments (OpenAI ProcGen), that the number of distinct game levels / scenes required for achieving perfect generalization far exceeds the number used by prior work. This points to the presence of overfitting in previous RL benchmarks and highlights the need for introducing more stochasticity in RL training environments.
- **Generalization Metric.** In contrast with supervised learning, the prior works in reinforcement learning measured both training and test performance on the same environment. The paper highlights the disadvantage of such an approach and presents a standard generalization metric to evaluate overfitting in reinforcement learning.
- **Factors affecting Generalization.** The paper evaluates the impact of different forms of regularization (*e.g.* stochastic policies, synthetic data augmentation, noisy environment dynamics) on the generalization performance in reinforcement learning.

4 Motivations from Computer Vision

4.1 SDC-Depth: Semantic Divide-and-Conquer Network for Monocular Depth Estimation.

Conference. CVPR 2020, Wang et al. (2020)

Summary. The paper tackles the problem of monocular depth estimation by incorporating scene priors from semantic and instance segmentation with a divide and conquer strategy. Wang et al. (2020) decompose the original image into multiple semantic and instance segments, which have very consistent depth structures and thus, are easier inputs for depth estimation. Finally, they propose a depth aggregation pipeline which combines depths maps at category and instance levels with a bottom-up approach (instance → category → global) to output global depth predictions.

Strengths.

- **Semantic Divide and Conquer**: The paper proposes a powerful idea of decomposing the overall depth estimation task for an image into its semantic constituents. Infact, the semantic divide and conquer strategy proposed in this paper forms a major motivation behind my recent work (Singh & Zheng, 2020) on using semantic guidance for *learning to paint*.

- **Novelity**: The paper brilliantly utilizes the consistency of depth structures in low level semantic/instance segments to propose an end to end training pipeline for monocular depth estimation.
- **Leveraging segmentation datasets**: Monocular depth estimation heavily relies on learning strong scene priors. Since, densely annotated depth datasets are limited, the proposed approach leverages large scale segmentation datasets to improve semantic understanding.
- **Ablation Studies**: The paper provides extensive ablation studies which highlight the importance of various model parts (Tab. 4) and the effect of segmentation data on depth prediction (Fig. 8).

5 Variance Reduction in Policy Gradients

5.1 High-Dimensional Continuous Control Using Generalized Advantage Estimation

Conference. ICLR, 2016.

Merits. While this paper is a bit old, it presents an essential mathematical analysis of variance reduction in policy gradient algorithms from the perspective of bias-variance tradeoff.

- **Bias-variance Tradeoff**: The paper presents a succinct mathematical formulation which allows for controlling the bias-variance tradeoff in advantage function estimation using a single hyperparameter (λ^{GAE}).
- **Interpretation as reward shaping**: The proposed formulation for advantage estimation can also be viewed as computing a low-bias estimate of the advantage function for an MDP with a *steeper* discount factor $\gamma' = \gamma\lambda$. (Note that the steeper discount factor $\gamma\lambda$ essentially removes variance introduced by all samples with delay greater than $1/(1-\gamma\lambda)$)
- **Ease of implementation**: The generalized advantage estimation method can be easily integrated with standard policy gradient algorithms like PPO, TRPO, A3C etc, using only few lines of code.

References

Cobbe, K., Klimov, O., Hesse, C., Kim, T., and Schulman, J. Quantifying generalization in reinforcement learning. *arXiv preprint arXiv:1812.02341*, 2018.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1126–1135. JMLR. org, 2017.

Packer, C., Gao, K., Kos, J., Krähenbühl, P., Koltun, V., and Song, D. Assessing generalization in deep reinforcement learning. *arXiv preprint arXiv:1810.12282*, 2018.

Singh, J. and Zheng, L. Combining semantic guidance and deep reinforcement learning for generating human level paintings. *arXiv preprint arXiv:2011.12589*, 2020.

Wang, L., Zhang, J., Wang, O., Lin, Z., and Lu, H. Sdc-depth: Semantic divide-and-conquer network for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 541–550, 2020.

Wortsman, M., Ehsani, K., Rastegari, M., Farhadi, A., and Mottaghi, R. Learning to learn how to learn: Self-adaptive visual navigation using meta-learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6750–6759, 2019.