Student ID : 20194293
Name : Go, Kyeong Ryeol

[AI 602] Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks

1. Paper Summary

Meta learning is the mechanism for learning to learn which aims to adapt to a particular task in a fast manner. The author proposed to train the meta-parameter which can be updated for the given task in a few gradient steps with a small amount of data. This is to build a general internal representation and to maximize the sensitivity of the loss functions with respect to the parameter. This is based on the intuition that some representation may be more transferrable than others.

The basic meta-learning problem is formulated as follow; during meta-training, a task T_i is sampled from the task distribution $p(T)$, the model is trained with K samples of the task, and then tested on new samples of the task. Model-Agnostic Meta-Learning (MAML) tries to find the parameter such that the gradient-based update rule can make rapid progress on new tasks drawn from $p(T)$ without overfitting. Therefore, the overall loss function can be summarized as below.

$$\min_{\theta} \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta_i'}) = \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta - \alpha \nabla_{\theta} L_{T_i}(f_{\theta})})$$

Here, one should notice that the optimization is conducted on the parameter θ while the objective is computed by the meta-parameter θ'. Computationally, a gradient through a gradient requires a Hessian-vector product. However, the first-order approximation which omits the second derivative actually performed comparably well with significant speed up suggesting that most of the improvement comes from the gradients of the objective at the post-update parameter values.

2. In-depth discussions

I. How can we get the theoretical guarantee for its theme of learning general internal representation? Is the visualization of the projection of the parameter space onto low dimensional space the only option?

II. How can we be more confident on its generalization performance on the new task? Are there any way to model the task relation which allows the unseen task to be represented as the linear combination of the previous seen tasks?