Paper-Reading-Notes $(universal)$


```
### MainTitle

- [x] **MainTitle** (conf) [[paperswithcode]()]
    - Author et al. "Title"


| 核心在哪? | 精读? 代码? | 关键词? | 亮点? | 笔记时间? |
| --------- | ----------- | ------- | ----- | --------- |
|           |             |         |       |           |

---



+ **背景? 提出了什么问题?**
+ **为了解决此问题提出了什么具体的idea?**
+ **如何从该idea形式化地对问题建模、简化并解决的?**
+ **理论方面证明的定理与推导过程?**
+ **这个任务/解决方法有什么意义?**
+ **对论文的讨论/感想?**
```

Adaptive Task Sampling (class-pair based)

- [x] **Adaptive Task Sampling (class-pair based)** (ECCV 2020) [[paperswithcode]()]
    - Liu et al. "Adaptive Task Sampling for Meta-Learning"

核心在哪?	精读? 代码?	关键词?	亮点?	笔记时间?

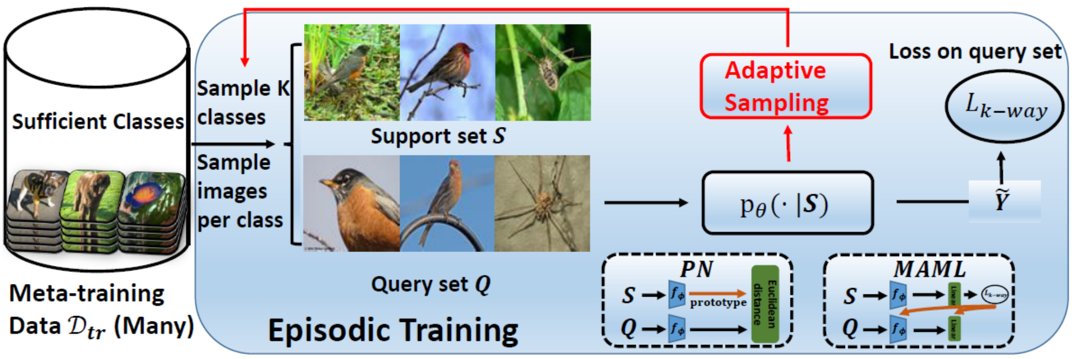

Fig. 1: The episodic training paradigm for meta-learning few-shot classification.

While a rich line of work focuses **solely on how to extract meta-knowledge across tasks**, we exploit the complementary problem on **how to generate informative tasks.**

We argue that the randomly sampled tasks could be **sub-optimal and uninformative** (e.g., the task of classifying “dog” from “laptop” is often trivial) to the meta-learner. In this paper, we propose **an adaptive task sampling method** to improve the generalization performance.

> In summary, our work makes the following contributions. **(1)** We propose a class-pair based adaptive task sampling approach for meta-learning methods, to improve the generalization performance on unseen tasks. **(2)** We further develop a greedy class-pair based approach that not only significantly reduces the complexity of task distribution computation, but also guarantees the generation of an identical distribution as that in the non-greedy approach. **(3)** We study the impact of the adaptive task sampling method by integrating it with various meta-learning approaches and performing comprehensive experiments on the miniImageNet and CIFAR-FS few-shot datasets, which quantitatively demonstrates the superior performance of our method. **(4)** We also conduct an extensive investigation of different sampling strategies, including class-based method, easy class-pair based method and uncertain class-pair based method. The results show that hard class-pair based sampling consistently leads to more accurate results.

- **背景？提出了什么问题？**

  an episodic training paradigm.

  A series of few-shot tasks are sampled from meta-training data for **the extraction of transferable knowledge across tasks**, which is then applied to **new** few-shot classification tasks consisting of unseen classes during the meta-testing phase.

  - 问题的提出：

    Despite their noticeable improvements, these meta-learning approaches leverage **uniform sampling** over classes to generate few-shot tasks, which ignores the intrinsic relationships between classes when forming episodes.

    上述方法是 uniform sampling，这忽略了forming episodes时候类之间的内在联系。在一些领域中 比如集成学习Adaboost对challenging training examples优先训练后续分类器。

    - **很自然的提出问题：** Can we perform adaptive task sampling and create more difficult tasks for meta-learning?
    - **难点：** one key challenge in task sampling is to define the difficulty of a task.

      > multiple classes. However, the difficulty of a class, and even the semantics of a class, is dependent on each other. For instance, the characteristics to discriminate “dog” from “laptop” or “car” are relatively easier to uncover than those for discriminating “dog” from “cat” or “tiger”. In other words, the difficulty of a task goes beyond the difficulty of individual classes, and adaptive task sampling should consider the intricate relationships between different classes.

  - **Review for Episodic Training:**

    1. In each episode of meta-training, we first **sample K classes** $\mathbb{L}^K \sim \mathbb{C}_{tr}$.
    2. Then, we **sample M and N labelled images per class** in $\mathbb{L}^K$ to **construct** the support set $\mathbb{S}=\{(s_m, y_m)_m\}$ and query set $\mathbb{Q}=\{(q_n, y_n)_n\}$, respectively.

       从之前sample的类里面sample出 $\mathbb{S}, \mathbb{Q}$.

3. The episodic training for few-shot learning 是在**query set**上最优, The model is parameterized by θ and the loss is the negative loglikelihood of the true class of each query sample, 即优化:

$$\ell(\theta)=\underset{(S,Q)}{\mathbb{E}}\left[-\sum_{(q_n,y_n)\in Q}\log p_\theta\left(y_n\mid q_n,S\right)\right]$$

$p_\theta\left(y_n\mid q_n,S\right)$ 是在support set上的分类概率.

注意啊上面的损失是在 **query set** 上测的, 但是训练(后验)是在**support**上的.

梯度下降 $\Delta\ell(\theta)$.

- **Review for Instance-base Adaptive Sampling for SGD:**

  Select Sample 的概率:

  - 第一次:

  $$p_0(i\mid\mathbb{D})=\frac{1}{|\mathbb{D}|}$$

  - 之后:

  instance i at iteration $t+1$ **according to the current prediction probability** $p\left(y_i\mid x_i\right)$ and the selection probability at previous iteration $p^t(i)$

  $$p^{t+1}(i)\propto\left(p^t(i)\right)^\tau e^{\alpha\left(1-p\left(y_i\mid x_i\right)\right)}$$

  where the hyperparameters τ is a discounting parameter and α scales the influence of current prediction.

  This multiplicative update method has a close relation to maximum loss minimization [47] and AdaBoost [16].

- 为了解决此问题提出了什么具体的**idea?**

  a straightforward extension of the instance-based sampling.

- 如何从该**idea**形式化地对问题建模、简化并解决的?

  - **Class-based Sampling**:

    We propose a class-based sampling (c-sampling) approach that **updates the class selection probability $p_C^{t+1}(c)$ in each episode.**

    具体选择类概率的更新公式如下:

    Given $\mathbb{S}^t$ and $\mathbb{Q}^t$ at episode t, we could update the class selection probability for each class in current episode $c\in\mathbb{L}_K^t$ in the following way,

    $$p_C^{t+1}(c)\propto\left(p^t(c)\right)^\tau e^{\alpha\frac{\sum_{(q_n,y_n)\in\mathbb{Q}^t}\mathbb{I}[c\neq y_n]p(c\mid q_n,\mathbb{S}^t)+\mathbb{I}[c=y_n](1-p(c\mid q_n,\mathbb{S}^t))}{NK}}. \tag{3}$$

    Note that we average the prediction probability of classifying each query sample n into incorrect classes in $\mathbb{L}_K^t$. Then we can sample K classes without replacement to construct the category set $\mathbb{L}_K^{t+1}$ for the next episode.

    每个类的难度不是独立的.

    取出类别二元组, 无向概率图模型 马尔可夫随机场, 这里不是最大团.

    更新$C(i,j)$, 该类别对在上一次就混淆了, 接下来就要挑这个.

    不能接受的计算复杂度, 则使用贪心算法.

- 理论方面证明的定理与推导过程**?**
- 这个任务**/**解决方法有什么意义**?**
- 对论文的讨论**/**感想**?**

Stochastic optimization with importance sampling for regularized loss minimization. (ICML 2015)