MELR: META-LEARNING VIA MODELING EPISODELEVEL RELATIONSHIPS FOR FEW-SHOT LEARNING

Nanyi Fei, Zhiwu Lu, Tao Xiang, Songfang Huang

ICLR 2021

In this paper

Address the cross-episode relation

Method:

1. Cross-Episode Attention Module
   a. To alleviate the effect of poor sampling (????)
2. Cross-Episode Consistency Regularization
   a. To enforce two classifier learned from two episodes consistent

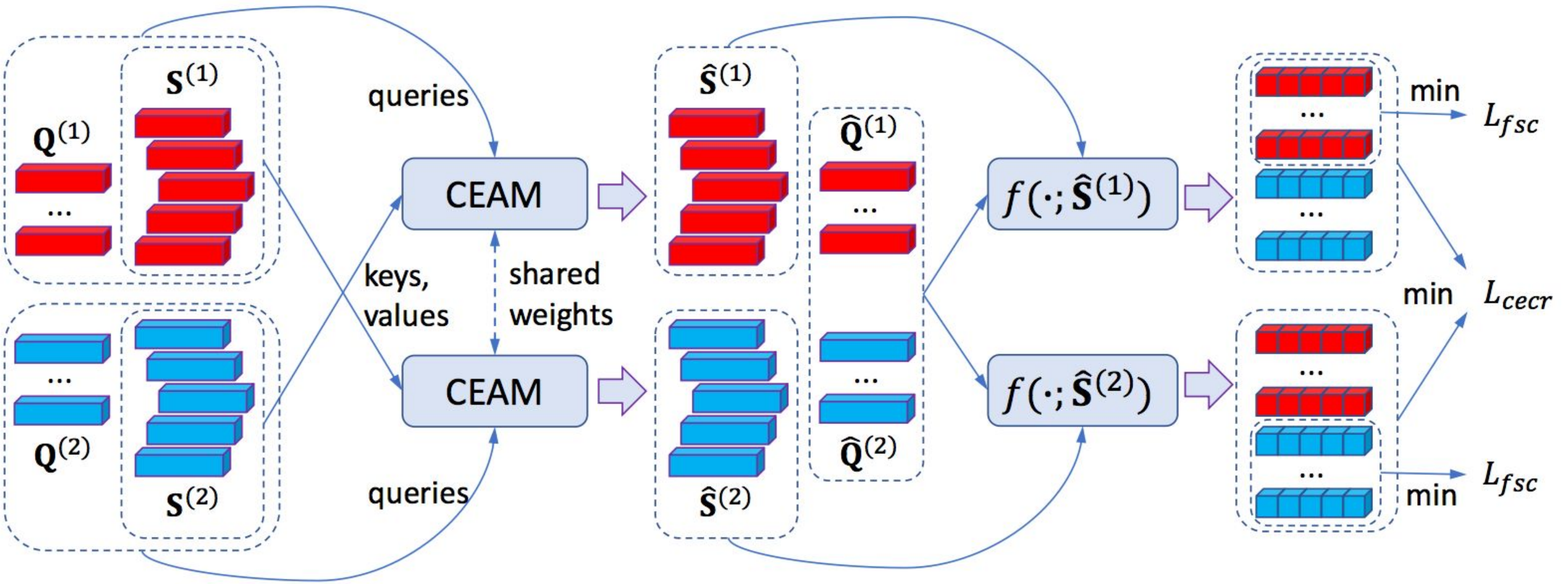
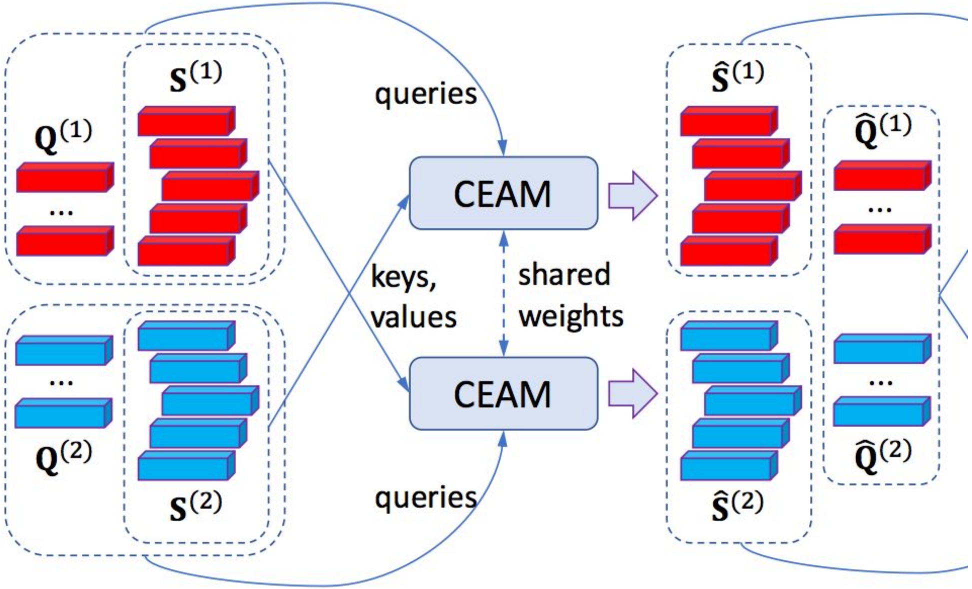


Figure 1: The schematic illustration of the proposed MELR model. It consists of two main components for modeling episode-level relationships: Cross-Episode Attention Module (CEAM) and Cross-Episode Consistency Regularization (CECR). For clarity, only the 5-way 1-shot setting is presented here. Each red/blue cuboid denotes a single instance.

Cross-episode attention



$$\hat{\mathbf{F}}^{(1)} = \mathrm{CEAM}(\mathbf{F}^{(1)}, \mathbf{S}^{(2)}, \mathbf{S}^{(2)}) = \mathbf{F}^{(1)} + \mathrm{softmax}(\frac{\mathbf{F}_Q^{(1)} \mathbf{S}_K^{(2)T}}{\sqrt{d}}) \mathbf{S}_V^{(2)},$$

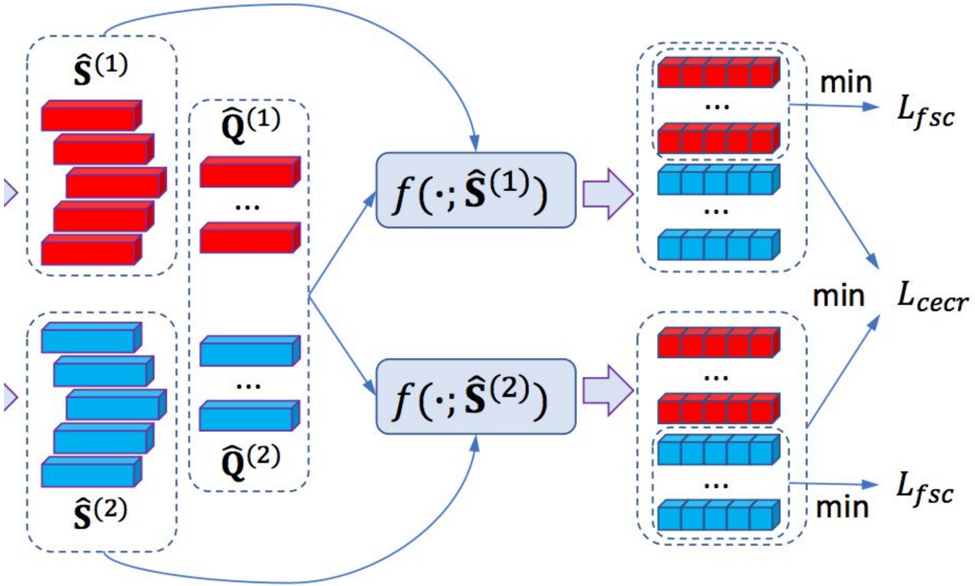
$$\hat{\mathbf{F}}^{(2)} = \mathrm{CEAM}(\mathbf{F}^{(2)}, \mathbf{S}^{(1)}, \mathbf{S}^{(1)}) = \mathbf{F}^{(2)} + \mathrm{softmax}(\frac{\mathbf{F}_Q^{(2)} \mathbf{S}_K^{(1)T}}{\sqrt{d}}) \mathbf{S}_V^{(1)},$$

Cross-episode Consistency

Train two classifiers

Enforce consistency

$$L'(f(\hat{\mathbf{q}}_i^{(1,2)};\hat{\mathbf{S}}^{(1)}), f(\hat{\mathbf{q}}_i^{(1,2)};\hat{\mathbf{S}}^{(2)});T)$$
$$= -\sum_{j=1}^{N} \sigma_j(f(\hat{\mathbf{q}}_i^{(1,2)};\hat{\mathbf{S}}^{(1)});T) \log\left(\sigma_j(f(\hat{\mathbf{q}}_i^{(1,2)};\hat{\mathbf{S}}^{(2)});T)\right).$$



Result

Method	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot
MatchingNet (Vinyals et al., 2016)	Conv4-64	43.56 ± 0.84	55.31 ± 0.73	–	–
ProtoNet† (Snell et al., 2017)	Conv4-64	52.78 ± 0.45	71.26 ± 0.36	53.82 ± 0.48	71.77 ± 0.41
MAML (Finn et al., 2017)	Conv4-64	48.70 ± 1.84	63.10 ± 0.92	51.67 ± 1.81	70.30 ± 0.08
RelationNet (Sung et al., 2018)	Conv4-64	50.40 ± 0.80	65.30 ± 0.70	54.48 ± 0.93	71.32 ± 0.78
IMP (Allen et al., 2019)	Conv4-64	49.60 ± 0.80	68.10 ± 0.80	–	–
DN4 (Li et al., 2019c)	Conv4-64	51.24 ± 0.74	71.02 ± 0.64	–	–
PARN (Wu et al., 2019)	Conv4-64	55.22 ± 0.84	71.55 ± 0.66	–	–
PN+rot (Gidaris et al., 2019)	Conv4-64	53.63 ± 0.43	71.70 ± 0.36	–	–
CC+rot (Gidaris et al., 2019)	Conv4-64	54.83 ± 0.43	71.86 ± 0.33	–	–
Centroid (Afrasiyabi Arman, 2020)	Conv4-64	53.14 ± 1.06	71.45 ± 0.72	–	–
Neg-Cosine (Liu et al., 2020)	Conv4-64	52.84 ± 0.76	70.41 ± 0.66	–	–
FEAT (Ye et al., 2020)	Conv4-64	55.15 ± 0.20	71.61 ± 0.16	–	–
MELR (ours)	Conv4-64	$\mathbf{55.35 \pm 0.43}$	$\mathbf{72.27 \pm 0.35}$	$\mathbf{56.38 \pm 0.48}$	$\mathbf{73.22 \pm 0.41}$

Mixup + Cross-Episode Consistency

Train a classifier on support set

Generate soft-label for unlabeled data

Mix soft-labeled data with golden data -> two episodes

Train two classifiers -> Enforce consistency based on query set