Boosting Few-Shot Learning with Adaptive Margin Loss

Aoxue Li, Weiran Huang, Xu Lan, Jiashi Feng, Zhenguo Li, Liwei Wang

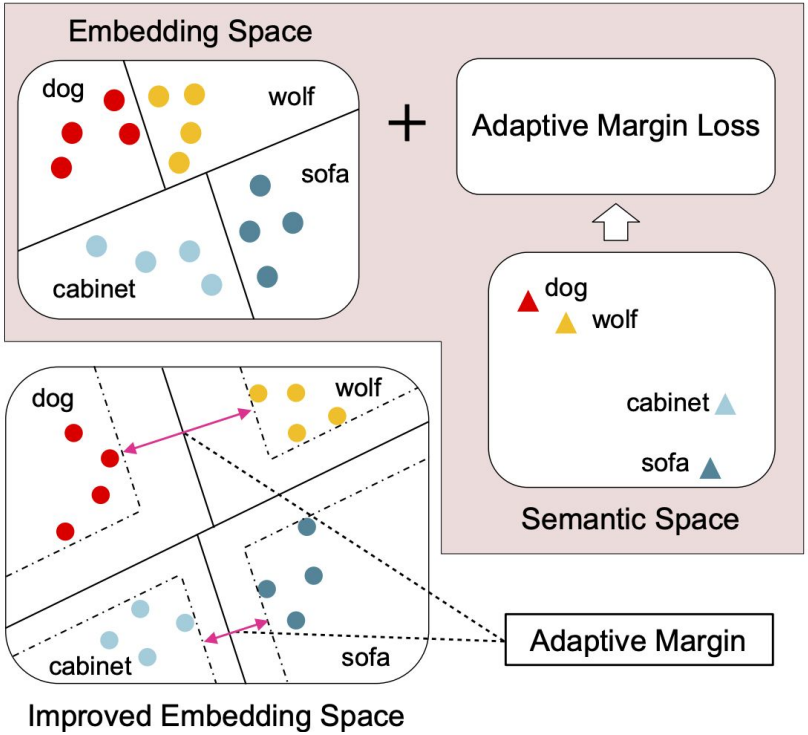
CVPR 2020

Intuition

Propose an adaptive margin principle to improve the generalization ability of metric-based meta-learning

Arguments:

Semantic similarity between different classes should be larger than the one between dissimilar classes



Naive Additive Margin Loss (NAML)

Increase the distances between classes

$$\mathcal{L}^{\mathrm{na}} = -\frac{1}{|Q|} \sum_{(x,y)\in Q} \log p^{\mathrm{na}}(y|x,S),$$

$$p^{\mathrm{na}}(y|x,S) = \frac{e^{\mathcal{D}(\mathcal{F}(x),r_y)}}{e^{\mathcal{D}(\mathcal{F}(x),r_y)} + \sum_{k\in C_t\setminus\{y\}} e^{\mathcal{D}(\mathcal{F}(x),r_k)+m}}.$$

Where: S, Q are support set and query set

F is an encoder function, r is class representation embedding

Class-Relevant Additive Margin Loss (CRAML)

Semantic similarity based on class name

$$m_{i,j}^{\mathrm{cr}} := \mathcal{M}(e_i, e_j) = \alpha \cdot \mathrm{sim}(e_i, e_j) + \beta,$$

Class-relevant additive margin loss

$$p^{\mathrm{cr}}(y|x,S) = \frac{e^{\mathcal{D}(\mathcal{F}(x), r_y)}}{e^{\mathcal{D}(\mathcal{F}(x), r_y)} + \sum_{k \in C_t \setminus \{y\}} e^{\mathcal{D}(\mathcal{F}(x), r_k)) + m_{y,k}^{\mathrm{cr}}}}$$

Task-Relevant Additive Margin Loss (TRAML)

$$\{m^{\mathrm{tr}}_{y,k}\}_{k\in C_t\setminus\{y\}} = \mathcal{G}\left(\{\mathrm{sim}(e_y, e_k)\}_{k\in C_t\setminus\{y\}}\right),$$

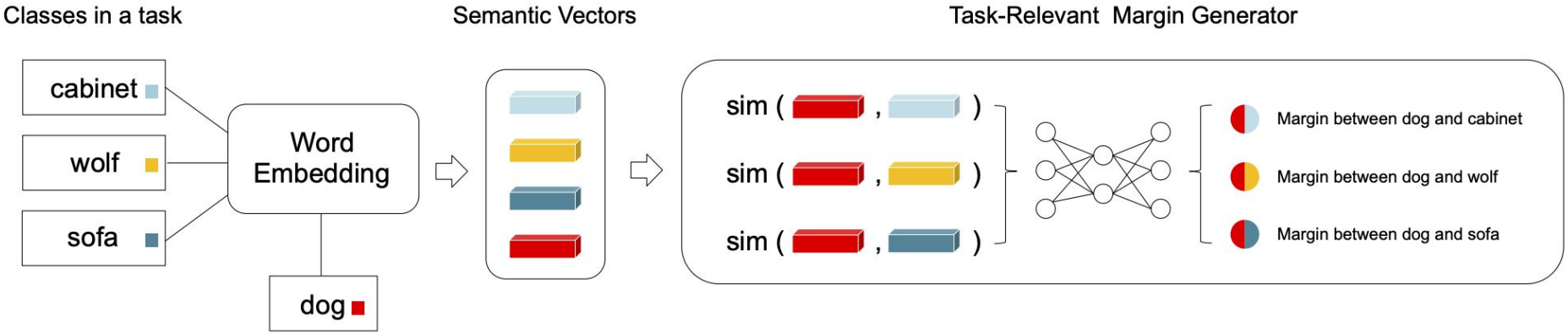


Figure 3. The illustration of the architecture of our task-relevant margin generator.

Result on miniImageNet

Model	Backbone	Type	Test Accuracy	
			5-way 1-shot	5-way 5-shot
Matching Networks [31]	4Conv	Metric	43.56 ± 0.84	55.31 ± 0.73
Prototypical Network [27]	4Conv	Metric	49.42 ± 0.78	68.20 ± 0.66
Relation Networks [27]	4Conv	Metric	50.44 ± 0.82	65.32 ± 0.70
GCR [15]	4Conv	Metric	53.21 ± 0.40	72.34 ± 0.32
Memory Matching Network [3]	4Conv	Metric	53.37 ± 0.48	66.97 ± 0.35
Dynamic FSL [8]	4Conv	Metric	56.20 ± 0.86	73.00 ± 0.64
Prototypical Network [27]	ResNet12	Metric	56.52 ± 0.45	74.28 ± 0.20
TADAM [20]	ResNet12	Metric	58.50 ± 0.30	76.70 ± 0.38
DC [17]	ResNet12	Metric	62.53 ± 0.19	78.95 ± 0.13
TapNet [36]	ResNet12	Metric	61.65 ± 0.15	76.36 ± 0.10
ECMSFMT [24]	ResNet12	Metric	59.00	77.46
AM3 (Prototypical Network) [35]	ResNet12	Metric	65.21±0.49	75.20 ± 0.36
MAML [7]	4Conv	Gradient	48.70 ± 1.84	63.11 ± 0.92
MAML++ [1]	4Conv	Gradient	52.15 ± 0.26	68.32 ± 0.44
iMAML [22]	4Conv	Gradient	49.30 ± 1.88	-
LCC [19]	4Conv	Gradient	54.6 ± 0.4	71.1 ± 0.4
CAML [11]	ResNet12	Gradient	59.23 ± 0.99	72.35 ± 0.18
MTL [28]	ResNet12	Gradient	61.20 ± 1.80	75.50 ± 0.80
MetaOptNet-SVM [12]	ResNet12	Gradient	62.64 ± 0.61	78.63 ± 0.46
Prototypical Network + TRAML (OURS)	ResNet12	Metric	60.31 ± 0.48	77.94 ± 0.57
AM3 (Prototypical Network) + TRAML (OURS)	ResNet12	Metric	**67.10** ± 0.52	**79.54** ± 0.60

Ablation Study on miniImagenet

Model	Test Accuracy	
	5-way 1-shot	5-way 5-shot
Original Classification Loss	65.21 ± 0.49	75.20 ± 0.36
Naive Additive Margin Loss	65.42 ± 0.25	75.48 ± 0.34
Class-Relevant Additive Margin Loss	66.36 ± 0.57	77.21 ± 0.48
Our Full Model	$\mathbf{67.10} \pm 0.52$	$\mathbf{79.54} \pm 0.60$