1

Meta learning for Semi-Supervised FSL

1. meta learning for semi-supervised FSL

K-shot N-way episodes.

total K example from N classes.

i) support (training) set $S = \{(x_1, y_1), (x_2, y_2) \cdots (x_{N\times K}, y_{N\times K})\}$

ii) Query/test set $Q = \{(x_1, y_1), (x_2, y_2) \cdots (x_T, y_T)\}$

Prototype, P_c of class c

$$P_c = \frac{h(x_i)\, z_{ic}}{\sum z_{ic}} \; ; \quad z_{ic} = 1_{[y_i = c]}$$

$$p(c \mid x^*, \{P_c\}) = \frac{\exp(-\|h(x^*) - P_c\|_2^2)}{\sum_{c'} \exp(-\|h(x^*) - P_{c'}\|_2^2)}$$

argmax $y^* = c$

Loss function: to minimize

$$-\frac{1}{T} \sum_i \log p(y_i^* \mid x_i^*, \{P_c\})$$

for test data (query)

Prototypical Net with soft K-means: extra terms [softmax]

$$\tilde{P}_c = \frac{\sum_i h(x_i)\, z_{i,c} + \sum_j h(\tilde{x}_j)\, \tilde{z}_{j,c}}{\sum_i z_{i,c} + \sum_j \tilde{z}_{j,c}} \; ; \quad \tilde{z}_{j,c} = \frac{\exp(-\|h(\tilde{x}_j) - P_c\|_2^2)}{\sum_{c'} \exp(-\|h(\tilde{x}_j) - P_{c'}\|_2^2)}$$

labeled

unlabeled

Refinement prototypes: $\tilde{P}_c$

Protop. Net with k-means [A distractor class]

Assumption: $p_c = \begin{cases} \frac{\sum_i h(x_i) z_{ic}}{\sum_i z_{ic}} ; & \text{for } c = 1 \dots N \\ 0 ; & \text{for } c = N+1 \end{cases}$

$$\tilde{z}_{j,c} = \frac{\exp\left(-\frac{1}{r_c^2} \| \tilde{x}_j - p_c \|_2^2 - A(r_c)\right)}{\sum_{c'} \exp\left(-\frac{1}{r_{c'}^2} \| \tilde{x}_j - p_{c'} \|_2^2 - A(r_{c'})\right)}$$

$h(\tilde{x}_j)$

length scale for distractor class.

; $A(r) = \frac{1}{2} \log(2\pi) + \log(r)$

$\boxed{\text{Here this paper, } r_{1,\dots N} = 1 \quad \text{learn } r_{N+1} = ??}$

PN soft k-means & masking:

Normalized Distance, $\tilde{d}_{j,c} = \frac{d_{jc}}{\frac{1}{M} \sum_j d_{jc}}$; $d_{jc} = \| h(x_j) - p_c \|_2^2$

soft threshold

$[\beta_c, \gamma_c] = MLP\left(\left[\min_j(\tilde{d}_{jc}), \max_j(\tilde{d}_{jc}), \operatorname{var}_j(\tilde{d}_{jc}), \operatorname{skew}_j(\tilde{d}_{jc}), \operatorname{kurt}_j(\tilde{d}_{jc})\right]\right)$

slope

$$\tilde{p}_c = \frac{\sum_i h(x_i) z_{ic} + \sum_j h(\tilde{x}_j) \tilde{z}_{jc} m_{jc}}{\sum_i z_{ic} + \sum_j \tilde{z}_{jc} m_{jc}} \; ; \quad m_{jc} = \sigma\left(-\gamma_c(\tilde{d}_{jc} - \beta_c)\right)$$

soft mask

a (modified term)

sigmoid

modified cluster.