⑰ Divide MIX

Divide Mix

Training Data $D = (X, Y) = \{(x_i, y_i)\}_{i=1}^{N}$

y_i → one hot.

Unlabed U

Cross entropy loss; $L(\theta) = \{L_i\}_{i=1}^{N} = \left\{ \sum_{c=1}^{C} y_i^c \log \left(p_{model}^c(x_i; \theta) \right) \right\}_{i=1}^{N}$

$p_{model}^c(x_i;\theta)$ → output for c

Penalize the entropy term:

$$fl = - \sum_{c} p_{model}^c(x;\theta) \log \left(p_{model}^c(x;\theta) \right)$$

// Avoid Near 0 noremalized loss.

minibatch labeled data / unlabeled data

$\{(x_b, y_b, w_b); b \in 1, \dots B\}$

$\{u_b; b \in (1, \dots B)\}$

gaussian component (smaller loss)

Data clean probability $w_i = p(g \mid L_i)$

loss value.

↓

compared with threshold τ

co-refinement: $\bar{y}_b = w_b y_b + (1 - w_b) p_b$

- y_b: provided label
- p_b: clean probability; Avg across different Augmentation (prediction by net)

Applying Sharpening:

$$\hat{y}_b = \text{sharpen}(\bar{y}_b, T) = \frac{\bar{y}_b^{c\,\frac{1}{T}}}{\sum_{c=1}^{C} \bar{y}_b^{c\,\frac{1}{T}}} \text{ for all } \forall c$$

Pair of sample (x_1, x_2) corresponding label (p_1, p_2)

mixed (x', p')

$$\lambda \sim \text{Beta}(\alpha, \alpha)$$
$$\lambda' = \max(\lambda, 1-\lambda)$$
$$x' = \lambda' x_1 + (1-\lambda') x_2$$
$$p' = \lambda' p_1 + (1-\lambda') p_2$$

Kind of Augmentation. (x' closer to x_1 than x_2)

Loss function:

{ Cross entropy loss } $\mathcal{L}_x = \frac{1}{|X'|} \sum_{x,p \in X'} \sum_c p_c \log\left(p_{model}^c(x;\theta)\right)$

{ consistency across net } $\mathcal{L}_u = \frac{1}{|U'|} \sum_{x,p \in U'} \| p - p_{model}(x,\theta) \|_2^2$

{ Regularization } $\mathcal{L}_{reg} = \sum_c \pi_c \log\left(\pi_c \Big/ \frac{1}{|X'|+|U'|} \sum_{x \in X'+U'} p_{model}^c(x,\theta)\right)$

(π_c: $1/c$ uniform)