Grow and merge.

Problem Definition:

two stage of CCD.

i) initial stage

ii) CCD stage.

training Dataset $D^0_{train} = \{(x_i^0, y_i^0)\}_{i=0}^{N^0}$

initial Known categories $C^0 = \{1, 2, \ldots K^0\}$

Serial of unlabeled dataset: $\{D^t_{train}\}_{t=0}^{T}$

$$D^t_{train} = \{x_i^t\}_{i=0}^{N^t}$$

$$C^t = \{1, 2, \ldots K^t\}$$

unknown categories

Evaluation metric:

$$D^t_{test} = \{(x_i^s, y_i^s) \mid s \leq t\}$$

// Examples from all previous classes.

maximum forgetting factor M_f

final Discovery metric M_d

Measures ACC^t_{known} & ACC^t_{novel}

where:

$$\boxed{M_d = ACC^T_{novel}}$$

$$\boxed{M_f = ACC^{0}_{known} - ACC^{t}_{known}}$$

Initial Stage:

Examplar Set $P_k = \{ p_{k,i} \}$

$$p_{k,i} = \arg\min_{x} \left\| \mu_k - \frac{1}{i} \left[\phi^0(x) + \sum_{j=1}^{i-1} \phi^0(p_{k,j}) \right] \right\|$$

CCD

Dual branch Architecture:

Static ϕ_S

Dynamic ϕ_D

initialized from ϕ^0

Growing

① Novelty detection

(v) Novel categories Discovery:

(i) Novelty Detection

Threshold based Novelty Detection.

$$d_{novel}(x_i^t) = \min_{1 \le k \le K^t} \underbrace{d(\mu_k, \phi_D(x_i^t))}_{\text{Distance function}}$$

x_i^t is novel if $\boxed{d_{novel}(x_i^t) \geq \epsilon}$

(ii) NCD & static branch

Optimize the $\phi_D^t(\cdot)$

(typical AutoNovel Approach)

(i)

$$\mathcal{L}_{BCE} = -\frac{1}{\hat{N}^t} \sum_{i=1}^{\hat{N}^t} \sum_{j=1}^{\hat{N}^t} \left(s_{ij} \log c_i^{t\,T} c_j^t + (1 - s_{ij}) \log(1 - c_i^{t\,T} c_j^t)\right)$$

Dynamic branch

winner take all (WTA) approach.

(2) Static - Dynamic Distillation

Static Branch

$$\mathcal{L}_{SD} = \frac{1}{N^t} \sum_{i=1}^{N^t} d\left(z_i^t, z_i'^t\right)$$

$$= \frac{1}{N^t} \sum_{i=1}^{N^t} d\left(\phi_S^t(x_i^t), \phi_D^t(x_i^t)\right)$$

~~Merging~~

(i) category unification: Continuous Learning for NCD.

1) Sample Sifting: filter incorrectly assigned samples.

Local Sample Density: $g_j(\hat{z}_i^t) = \max_{z \in N_j(\hat{z}_i^t)} d\left(\hat{z}_i^t, z\right)$

$N_j(\hat{z}_i^t)$ j-th nearest sample.

Sample with higher ξ_j are sifted out.

ii) Pseudo-label Representation Learning:

$$\mathcal{L}_{PLL} = -\frac{1}{K^t}\sum_{k=1}^{K^t}\frac{1}{|P_k|}\sum_{i=1}^{|P_k|}\log\frac{\exp\left(\phi_D^t(P_{k,i})\cdot \mu_k/\tau\right)}{\sum_{j=1}^{k^t}\exp\left(\phi_D^t(P_{k,i})\cdot \mu_j/\tau\right)}$$

Branch Unification.

$$\boxed{\theta_{\phi_D^t} \leftarrow \alpha\,\theta_{\phi_S} + (1-\alpha)\,\theta_{\phi_D^t}}$$

$\rightarrow$ momentum encoder. setting.