Neighborhood contrastive learning.

Problem statement:

given: labeled set D^l

unlabeled D^u

target: learn to cluster D^u, leveraging D^l

feature extractor $\Omega : x \rightarrow z \in \mathbb{R}^H$

$\phi^l \rightarrow c^l$

$\phi^u \rightarrow c^u$

Baseline:

$(x, y) \in D^l$

optimize cross-entropy loss:

$$\boxed{l_{CE} = -\frac{1}{c^l} \sum_{i=1}^{c^l} y_i \log \phi_i^l(\Omega(x))}$$

pairwise ~~binary~~ CE loss:

$$x_i^u, \; x_j^u \in \mathcal{D}^u \xrightarrow{f} z_i^u, \; z_j^u$$

$$\text{cosSim}: \; \delta(z_i^u, z_j^u) = \frac{z_i^{u\top} \cdot z_j^u}{\|z_i^u\|_2^2 \, \|z_j^u\|_2^2}$$

$$\hat{y}_{ij} = \mathbb{1}\left[\delta(z_i^u, z_j^u) \geq \lambda\right]$$

$$l_{bce} = \hat{y}_{ij} \log p_{ij} + (1 - \hat{y}_{ij}) \log(1 - p_{ij})$$

$$p_{ij} = \underline{\underline{\phi}}^u(z_i^u)^\top \, \underline{\underline{\phi}}^u(z_j^u)$$

Extra head compared to prev. work.

Consistency loss:

$$l_{mse} = \frac{1}{C^l} \sum_{i=1}^{C^l} \left(\phi_i^l(z^l) - \phi_i^l(\hat{z}^l)\right)^2 + \frac{1}{C^u} \sum_{j=1}^{C^u} \left(\phi_j^u(z^u) - \phi_j^u(\hat{z}^u)\right)^2$$

$$l = l \quad l \quad + w(t)\, l$$

$$\boxed{\mathcal{L}_{uace} = \ell_{CE} + \lambda_{CE} \cdot \ell_{MSE}}$$

Neighborhood Contrastive L:

Queue : M^u [regarded as negative]

sample $\bar{z}^u$

$$\ell(z^u, \hat{z}^u) = -\log \frac{e^{\delta(z^u, \hat{z}^u)/\tau}}{e^{\delta(z^u, \hat{z}^u)/\tau} + \underbrace{\sum_{m=1}^{|M^u|} e^{\delta(z^u, \bar{z}^u)/\tau}}_{\text{fake negative issue.}}}$$

Retrieve top-k similar features:

$$\rho_k = \operatorname*{argtop_k}_{\bar{z}_i^u} \left(\left\{ \delta(z^u, \bar{z}_i^u) \;\middle|\; \forall i \in \{1 \cdots |M^u|\} \right\} \right)$$

consider them as pseudo positive

$$\ell(z^u, \rho_k) = -\frac{1}{K} \sum_{\bar{z}^u \in \rho} \log \frac{e^{\delta(z^u, \bar{z}_i^u)/\tau}}{\delta(z^u, \bar{z}^u)/\tau + \sum^{|M^u|} e^{\delta(z^u, \bar{z}_m^u)/\tau}}$$

$$\ldots \quad e \quad \ldots \sum_{m=1}$$

$\therefore$ NCL :

$$\boxed{l_{NCE} = \alpha\, l_{(z^u,\, \hat{z}^u)} + (1-\alpha)\, l_{(z^u,\, r_k)}}$$

Supervised contrastive loss:

$$P = \{ \tilde{z}_j^l \in M^l : y_i = y_j \} \cup \hat{z}_i^l$$

$$\boxed{l_{scl} = -\frac{1}{|P|} \sum_{\tilde{z}_j^l \in P} \log \frac{e^{\delta(z_i^l,\, \tilde{z}_j^l)/\tau}}{e^{\delta(z_i^l,\, \hat{z}_i^l)} + \sum_{m=1}^{|M^l|} e^{\delta(z_i^l,\, \tilde{z}_m^l)/\tau}}}$$

Hard negative generation:

Easy negatives:

$$\epsilon_k = \underset{\tilde{z}_i^u}{\arg\operatorname{top}} \left(\{ -\delta(z^u, \tilde{z}_i^u) \mid \forall i \in 1 \ldots |M^u| \} \right)$$

for samples in M^l, true negative to M^u

$\in E_k$

$$\delta = \mu \cdot \bar{z}^u + (1-\mu)\, \bar{z}^l$$

mixing coef.

randomly sampled

Hard negative:

$$\eta_k = \underset{\delta_i}{\arg\operatorname{top}_k} \left(\left\{ \delta(z^u, \delta_i) \,\middle|\, \forall i \in \{1, \dots, N\} \right\} \right)$$

New memory queue:

$$M^{u'} = M^u \cup \eta_k$$

Overall loss:

$$Loss = L_{base} + L_{scl} + L_{rel}$$