labeled data $D^l = \{x^l, y^l\} \rightarrow 1 \dots N^l$

unlabeled $D^u = \{x^u\} \, 1 \dots N^u$

$$y_i^l \in \{1, \dots C^l\}$$

$D^l, D^u \rightarrow$ Disjoint classes.

get knowledge from D^l

transfer to discover

Naive Baseline :

① Model initialization

one hot of y_i^l

$$l_{CE} = -\frac{1}{n^l} \sum_{i=1}^{n^l} \log \left[\text{softmax} \, (z_i^l) \right]^T \cdot \hat{y}_i^l$$

$z_i^l \rightarrow$ output of classifier.

(ii) Unsupervised clustering

- pseudo-pair learning : (PPL)

cosine similarity matrix : $S \in \mathbb{R}^{n^u} \times \mathbb{R}^{n^u}$

$$S_{i,j} = \frac{(\hat{z}_i^u)^T \cdot \hat{z}_j^u}{\|\hat{z}_i^u\|_2^2 \; \|\hat{z}_j\|_2^2}$$

requires c^u info.

$\in \mathbb{R}^{c^u}$

$$\hat{z}_i^u = \mathrm{softmax}(z_i^u)$$

Based on th.

$$W_{i,j} = \begin{cases} 0 & ; \; s_{ij} < \theta_1 \\ 1 & ; \; s_{ij} > \theta_1 \end{cases}$$

$$\mathcal{L}_{PPL} = - \frac{1}{(n^u)^2} \sum_{ij} \Big(W_{ij} \log s_{i,j} + (1 - w_{ij}) \log (1 - s_{i,j}) \Big)$$

Bolstering belief in expectation should be good.

dynamic changes.

pseudo-label learning :

- Pseudo-label reweighting.

One hot pseudo-label:

one issue if there is distributed representation

$$\hat{y}_i^u[j] = \begin{cases} 0 & , \hat{z}_i^u[j] < \theta_2 \\ 1 & ; \hat{z}_i^u[j] > \theta_1 \end{cases}$$

→ possible solution Entropy [my proposal]

$$\mathcal{L}_{PLL} = -\frac{1}{n^u} \sum_i \log(\hat{z}_i^u)^T \cdot \hat{y}_i^u \;; \quad \forall i \in \max(\hat{y}_{ij}) = 1$$

final loss: $\boxed{\mathcal{L}_{UC} = \mathcal{L}_{PPL} + \lambda_i \mathcal{L}_{PLL}}$

OpenMix:

① Unmix with labeled data:

mix label + U data & labels

$$\lambda \sim \text{Beta}(\epsilon, \epsilon)$$ → hyperparam

$$\lambda^* \sim \max(\lambda, 1-\lambda) \geq 0.5$$

data; $m = \lambda^* x^l + (1-\lambda^*) x^u$

label: $\underline{v} = \lambda^* \bar{y}^l + (1-\lambda^*) \bar{y}^u$

closer to $\bar{x}^l$

[Helps avoiding overfitting]

to sample $[\max(\hat{\epsilon}^u) > \theta_2]$

(ii) **mix with reliable anchor:**

mix the anchor with the unlabeled data.

Hierarchical approach

RMSE loss ?? not classifier.

Loss of OpenMix

$$\mathcal{L}_{opm} = \frac{1}{|M|} \sum_{i \in M} \frac{1}{c^l + c^u} \| \underline{v_i} - \text{softmax}(\hat{\epsilon}_i^m) \|_2$$

overall loss:

$$\mathcal{L}_{all} = \mathcal{L}_{uc} + \lambda \mathcal{L}_{opm}$$

refers to transformed

Final loss:

Final Loss:

image.

$$\mathcal{L}_{all} = \mathcal{L}_{uc} + \lambda_2 \mathcal{L}_{opm} + \mathcal{L}'_{uc} + \lambda_2 \mathcal{L}'_{opm}$$