Multi-similarity with GPW for deep metric learning

GPW

$x_i \in \mathbb{R}^d$ // instance vector

$X \in \mathbb{R}^{m \times d}$ // instance matrix

$y \in \{1, 2 - - c\}^m$ //label vector

x_i Projected to unit sphere by $f(x_i; \theta)$

similarity $S_{ij} := \langle f(x_i; \theta), f(x_j; \theta) \rangle$

Pair based loss $\mathcal{L}(S, y)$

Derivative

$$\frac{\partial L}{\partial \theta}\Big|_t = \frac{\partial \mathcal{L}(s,y)}{\partial s}\Big|_t \frac{\partial s}{\partial \theta}\Big|_t$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial L(s,y)}{\partial s_{ij}}\Big|_t \frac{\partial f_{ij}}{\partial \theta}\Big|_t$$

Formulating

$$f(s,y) = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial \mathcal{L}(s,y)}{\partial s_{ij}}\Big|_t s_{ij}$$

$$\frac{\partial \mathcal{L}(s,y)}{\partial s_{ij}} \geq 0 \quad \text{// negative pair}$$

$$\leq 0 \quad \text{// positive pair}$$

$$\boxed{\begin{array}{l} \text{[increasing } s_{ij} \gg \text{loss decrease for +ve} \\ \text{increasing } s_{ij} \gg \text{loss increase -ve]} \end{array}}$$

So,
$$f = \sum_{i=1}^{m}\left(\sum_{y_j \neq y_i}^{m} \frac{\partial \mathcal{L}}{\partial s_{ij}}\Big|_t s_{ij} + \sum_{y_i = y_j}^{m} \frac{\partial \mathcal{L}}{\partial s_{ij}}\Big| s_{ij} \right)$$

$$= \sum_{i=1}^{m} \left(\sum_{y_j \neq y_i}^{m} w_{ij} s_{ij} - \sum_{y_j = y_i}^{m} w_{ij} s_{ij} \right)$$

$$\left| \frac{\partial \mathcal{L}(S, y)}{\partial s_{ij}} \right|_t$$

Revisiting paire-based loss

1. contrastive loss

$$\mathcal{L}_{cont} := (1 - I_{ij}) \underbrace{[s_{ij} - \lambda]_+}_{} - I_{ij} s_{ij}$$

indicator 0 // -ve pair
1 // +ve "

similar weight for all pairs $> \lambda$

2. Triplet loss

$$\mathcal{L}_{trip} := [S_{an} - S_{ap} + \lambda]_+$$

3. Lifted structure loss

$$\mathcal{L}_{lifted} = \sum_{i=1}^{m} \left[\log \sum_{y_k = y_i} e^{\lambda - S_{ik}} + \log \sum_{y_k \neq y_i} e^{S_{ik}}\right]_+$$

generalized +ve approach.

Here,

$$w^{+}_{ij} = \frac{e^{\lambda - S_{ij}}}{\sum_{y_k = y_i} e^{+\lambda - S_{ij}}} = \frac{1}{\sum_{y_k = y_i} e^{S_{ij} - S_{ik}}}$$

$$w^{-}_{ij} = \frac{e^{S_{ij}}}{\sum_{y_k \neq y_i} e^{S_{kj}}} = \frac{1}{\sum_{y_k \neq y_i} e^{-S_{ij} + S_{ik}}}$$

compared for relative Similarity

4. Binomial deviance loss: softplus function instead of hinge

$$\mathcal{L}_{bi} = \sum_{i=1}^{m} \left\{ \frac{1}{P_i} \sum_{y_i = y_j} \log\left[1 + e^{\alpha(\lambda - S_{ij})}\right] + \frac{1}{N_i} \sum_{y_i \neq y_j} \log\left[1 + e^{\beta(S_{ij} - \lambda)}\right] \right.$$

Proposed MS loss

$$\mathcal{L}_{MS} = \frac{1}{m} \sum_{i=1}^{m} \left\{ \frac{1}{\alpha} \log\left[1 + \sum_{k \in P_i} e^{-\alpha(S_{ik} - \lambda)}\right] + \frac{1}{\beta} \log\left[1 + \sum_{k \in N_i} e^{+\beta(S_{ik} - \lambda)}\right] \right\}$$

Gradient descent to Optimize.