Deep Hashing with Active Pairwise Supervision

Ziwei Wang[1,2,3], Quan Zheng[1,2,3], Jiwen Lu[1,2,3] ⋆, and Jie Zhou[1,2,3,4]

[1] Department of Automation, Tsinghua University, China
[2] State Key Lab of Intelligent Technologies and Systems, China
[3] Beijing National Research Center for Information Science and Technology, China
[4] Tsinghua Shenzhen International Graduate School, Tsinghua University, China
{wang-zw18,zhengq16}@mails.tsinghua.edu.cn,
{lujiwen,jzhou}@tsinghua.edu.cn

Supplementary

A. Proof of (5)

According to (4), we obtain the following inequality for the empirical risk:

$$\mathbb{E}_M(J) \leqslant \hat{\mathbb{E}}_M(J) + \Phi \tag{1}$$

Where $\Phi = 2R_c(\Omega) + \sqrt{\frac{\ln 1/\delta}{c}}$ means the model complexity. Using the true risk $E(J)$ to minus the above inequality, we can obtain (5).

B. Proof of Three Properties of the Distance Defined in (10)

We define the distance between binary code pairs as follows:

$$\begin{aligned} &d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{t})) \\ =&\min(||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_b)||_F, \\ &||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_b)||_F) \end{aligned}$$

Generally, the distance in the Hamming space has three fundamental mathematical properties: non-negativity, symmetry and triangle inequality. We show that the defined distance satisfies the above three properties in the following.

Non-negativity and Symmetry:
These two properties are obvious according to the non-negativity and symmetry of the Frobenius norm.

Triangle Inequality:
The triangle inequality means that for any sample pairs $\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{s}$, we have

$$d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{t})) + d(\mathcal{H}(\boldsymbol{t}), \mathcal{H}(\boldsymbol{s})) \geq \ d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{s}))$$

⋆ Corresponding author

We prove it in the following:

$$\begin{aligned}
& d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{t})) + d(\mathcal{H}(\boldsymbol{t}), \mathcal{H}(\boldsymbol{s})) \\
= & \min\Big\{ ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_b)||_F, \\
& ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_b)||_F \Big\} + \\
& \min\Big\{ ||\mathcal{H}(\boldsymbol{t}_a) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{t}_b) - \mathcal{H}(\boldsymbol{s}_b)||_F, \\
& ||\mathcal{H}(\boldsymbol{t}_b) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{t}_a) - \mathcal{H}(\boldsymbol{s}_b)||_F \Big\} \\
= & \min\Big\{ A, B \Big\} + \min\Big\{ C, D \Big\} \\
= & \min\Big\{ A + C, B + C, A + D, B + D \Big\}
\end{aligned}$$

Where

$$\begin{aligned}
& A + C \\
= & \Big\{ ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_b)||_F \Big\} + \\
& \Big\{ ||\mathcal{H}(\boldsymbol{t}_a) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{t}_b) - \mathcal{H}(\boldsymbol{s}_b)||_F \Big\} \\
= & \Big\{ ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{t}_a)||_F + ||\mathcal{H}(\boldsymbol{t}_a) - \mathcal{H}(\boldsymbol{s}_a)||_F \Big\} + \\
& \Big\{ ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{t}_b)||_F + ||\mathcal{H}(\boldsymbol{t}_b) - \mathcal{H}(\boldsymbol{s}_b)||_F \Big\} \\
\geq & ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{s}_b)||_F \\
\geq & \min\Big\{ ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{s}_b)||_F, \\
& ||\mathcal{H}(\boldsymbol{x}_b) - \mathcal{H}(\boldsymbol{s}_a)||_F + ||\mathcal{H}(\boldsymbol{x}_a) - \mathcal{H}(\boldsymbol{s}_b)||_F \Big\} \\
= & \ d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{s}))
\end{aligned}$$

Similarly, we have

$$B + C, A + D, B + D \geq d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{s}))$$

Thus,

$$\begin{aligned}
& d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{t})) + d(\mathcal{H}(\boldsymbol{t}), \mathcal{H}(\boldsymbol{s})) \\
= & \min\Big\{ A + C, B + C, A + D, B + D \Big\} \\
\geq & \ d(\mathcal{H}(\boldsymbol{x}), \mathcal{H}(\boldsymbol{s}))
\end{aligned}$$

Q.E.D.

C. Proof of (11)

Similar to [1], we hope to minimize the MMD objective:

$$MMD[\mathcal{L} \cup \mathcal{Q}, \mathcal{U} \setminus \mathcal{Q}] = \inf_{\boldsymbol{k}_1, \boldsymbol{k}_2} \Big\| \frac{1}{l+q} \sum_{i=1}^{l+q} \mathcal{T}_{k_{1,i}}\big(\mathcal{H}(\boldsymbol{x}_{1,i})\big) - \frac{1}{u-q} \sum_{i=1}^{u-q} \mathcal{T}_{k_{2,i}}\big(\mathcal{H}(\boldsymbol{x}_{2,i})\big) \Big\|_F^2 \tag{2}$$

where $\boldsymbol{x}_{1,i} \in \mathcal{L} \cup \mathcal{Q}$ is the i_{th} pair sampled from the labeled and query sets, and $\boldsymbol{x}_{2,i} \in \mathcal{U} \setminus \mathcal{Q}$ is the i_{th} pair sampled from the unlabeled excluding query instances. $k_{1,i}$ and $k_{2,i}$ is the i_{th} element of the permutation indicator $\boldsymbol{k}_1 \in \{0,1\}^{l+q}$ and $\boldsymbol{k}_2 \in \{0,1\}^{u-q}$.

Next we define a binary vector $\boldsymbol{\alpha}$ of size u to demonstrate the sample selection. Thus the problem reduces to finding $\boldsymbol{\alpha}$ that minimize the MMD objective:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \Big\| & \frac{1}{l+q}\big(\sum_{i=1}^{l} \mathcal{T}_{k_{L,i}}(\mathcal{H}(\boldsymbol{x}_{L,i})) + \sum_{j=1}^{u} \alpha_j \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j}))\big) \\ & - \frac{1}{u-q} \sum_{j=1}^{u} (1-\alpha_j) \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j})) \Big\|_F^2 \\ & s.t. \quad \boldsymbol{\alpha} \in \{0,1\}^u, ||\boldsymbol{\alpha}||_1 = q \end{aligned} \tag{3}$$

where we define $\boldsymbol{x}_{U,i}$ and $\boldsymbol{x}_{L,i}$ as the i_{th} pair sampled from the unlabeled and labeled sets. $k_L \in \{0,1\}^l$ and $k_U \in \{0,1\}^u$ as the corresponding permutation indicator, whose j_{th} elements are denoted as $k_{L,j}$ and $k_{U,j}$ respectively. α_j represents for the j_{th} element of $\boldsymbol{\alpha}$.

We can rewrite (2) as:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \Big\| & \sum_{i=1}^{l} \mathcal{T}_{k_{L,i}}(\mathcal{H}(\boldsymbol{x}_{L,i})) + \frac{n}{u-q} \sum_{j=1}^{u} \alpha_j \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j})) \\ & - \frac{l+q}{u-q} \sum_{j=1}^{u} \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j})) \Big\|_F^2 \end{aligned}$$

This is equivalent to:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \Big\| & \sum_{j=1}^{u} \alpha_j \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j})) + \frac{u-q}{n} \sum_{i=1}^{l} \mathcal{T}_{k_{L,i}}(\mathcal{H}(\boldsymbol{x}_{L,i})) \\ & - \frac{l+q}{n} \sum_{j=1}^{u} \mathcal{T}_{k_{U,j}}(\mathcal{H}(\boldsymbol{x}_{U,j})) \Big\|_F^2 \end{aligned}$$

Finally we rewrite the above equation and obtain the learning objective:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{K}_{UU}\boldsymbol{\alpha} + \frac{u-q}{n}\mathbf{1}^l \boldsymbol{K}_{LU}\boldsymbol{\alpha} - \frac{l+q}{n}\mathbf{1}^u \boldsymbol{K}_{UU}\boldsymbol{\alpha} \\ s.t. \quad & \boldsymbol{\alpha} \in \{0,1\}^u, ||\boldsymbol{\alpha}||_1 = q \end{aligned} \tag{4}$$

where the element in the i_{th} row and j_{th} column of $\boldsymbol{K}_{UU}$ and $\boldsymbol{K}_{LU}$ is denoted as $K_{UU,ij}$ and $K_{LU,ij}$ respectively:

$$\begin{aligned} K_{UU,ij} &= \inf_k \mathcal{H}(\boldsymbol{x}_{U,i})^T \mathcal{T}_k(\mathcal{H}(\boldsymbol{x}_{U,j})) \\ K_{LU,ij} &= \inf_k \mathcal{H}(\boldsymbol{x}_{L,i})^T \mathcal{T}_k(\mathcal{H}(\boldsymbol{x}_{U,j})) \end{aligned}$$

References

1. Chattopadhyay, R., Wang, Z., Fan, W., Davidson, I., Panchanathan, S., Ye, J.: Batch mode active sampling based on marginal probability distribution matching. TKDD **7**(3), 13 (2013)