1 Introduction

1.1 ALECH

Adaptive Label correlation based asymmEtric Cross-modal Hashing method, i.e., ALECH, is proposed for cross-modal retrieval[1].Other than most existing approaches, ALECH combines the adaptive label correlation and pairwise semantic similarity constraints. Consequently, the binary orthogonal optimization problem is as follows:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{B},\mathbf{Z}} ||\mathbf{PB}-\mathbf{ZL}||_F^2 + ||\mathbf{L}-\mathbf{ZL}||_F^2 + \alpha||\mathbf{B}^T\mathbf{B}-r\mathbf{S}||_F^2, \\ &s.t. \quad \mathbf{B} \in \{-1,1\}^{r\times n}, \mathbf{BB}^T = n\mathbf{I}_r, \mathbf{B1}_n = \mathbf{0}_r. \end{aligned} \tag{1}$$

ALECH decomposes hash learning into two steps, hash codes learning and hash functions learning.After obtaining the optimal hash codes $\mathbf{B}$, ALECH propose this hash functions learning framework:

$$\min_{\mathbf{W}^{(k)}} \lambda \left\|\mathbf{B}-\mathbf{W}^{(k)}\mathbf{X}^{(k)}\right\|_F^2 + \left\|r\mathbf{S}-\mathbf{B}^T\left(\mathbf{W}^{(k)}\mathbf{X}^{(k)}\right)\right\|_F^2 \tag{2}$$

1.2 MGBO

Nonmonotone Manifold Gradient method for Binary Orthogonal problem, i.e., MGBO, is featured with a restricted Stiefel manifold and a matrix box set[2]. Its penalty problems induced by the ℓ_1-distance from the box set and its Moreau envelope are both proved to be well-defined and served as the global exact penalties. Moreover, The penalty problem induced by the Moreau envelope is a smooth optimization which can be solved by a retraction-based line-search Riemannian gradient method.

Algorithm 1 (**N**onmonotone **M**anifold **G**radient method for **B**inary **O**rthogonal problem (**MGBO**))

Initialization: Select $\rho > 0$, $\gamma > 0$. Choose $\varepsilon > 0$, $\eta \in (0,1)$, $0 < \alpha < 1$, $0 < t_{\min} \le t_{\max}$ and $X^0 \in \mathcal{B}_v(n,r)$.
for $k = 0, 1, 2, \cdots$
Choose $t_k \in [t_{\min}, t_{\max}]$;
Let $V^k = -t_k \operatorname{grad}\Theta_{\rho,\gamma}(X^k)$ and $X^{k+1} = R_{X^k}(V^k)$;
if $\|\operatorname{grad}\Theta_{\rho,\gamma}(X^{k+1})\| \le \varepsilon$, return; **end if**.
Let $J_k = \{\max\{0, k-m\}, \ldots, k\}$;
Set $\ell = 0$, $t_{k,\ell} = t_k$, and $V^{k,\ell} = V^k$.
while $\Theta_{\rho,\gamma}(X^{k+1}) > \max\limits_{j\in J_k} \Theta_{\rho,\gamma}(X^j) - \frac{\alpha}{2t_{k,\ell}}\|V^{k,\ell}\|_F^2$ **do**
  Let $t_{k,\ell+1} = \eta t_{k,\ell}$;
  Set $V^{k,\ell+1} = -t_{k,\ell+1} \operatorname{grad}\Theta_{\rho,\gamma}(X^k)$;
  Set $X^{k+1} = R_{X^k}(V^{k,\ell+1})$;
  end while
end for

2 Improved ALECH by MGBO

2.1 Update B by MGBO

MGBO has equipped us with a powerful tool for solving the Binary Orthogonal Problem. By setting the derivatives of the objective function with respect to matrices $\mathbf{P}$ and $\mathbf{V}$ to zero, we can derive their closed-form solutions for Problem (1). This theoretical foundation enables us to develop Algorithm 2, which employs MGBO to iteratively update the binary code matrix $\mathbf{B}$.

Algorithm 2 (Improved ALECH by MGBO — Update $\mathbf{B}$ by MGBO)

Input: Training data $\{\mathbf{X}^{(k)}\}_{k=1}^{v}$, label matrix $\mathbf{L}$, code length r, model parameters α, β, λ, and maximum iteration number τ.

Output: Hash codes $\mathbf{B}$ and hash functions $\{\mathbf{W}^{(k)}\}_{k=1}^{v}$.

1: **Initialization:** Randomly initialize $\mathbf{B}$ with a standard normal distribution, $\mathbf{Z} = \mathbf{I}$;

2: Compute $\mathbf{G}$ by $\mathbf{G} = \left[\frac{\mathbf{L}_{*1}}{\|\mathbf{L}_{*1}\|_2}, \frac{\mathbf{L}_{*2}}{\|\mathbf{L}_{*2}\|_2}, \ldots, \frac{\mathbf{L}_{*n}}{\|\mathbf{L}_{*n}\|_2}\right]$;

3: **for** $i = 1$ to τ **do**

4: Update $\mathbf{P}$ by $\mathbf{P} = \mathbf{ZLB}^T/n$;

5: Update $\mathbf{Z}$ by $\mathbf{Z} = (\mathbf{PBL}^T + \mathbf{LL}^T)(2\mathbf{LL}^T)^{-1}$;

6: Update $\mathbf{B}$ by MGBO;

7: **end for**

8: **return** Hash codes $\mathbf{B}$.

9: Map the training data to kernel space by $\phi(\mathbf{x}) = \left[\exp\left(\frac{-\|\mathbf{x}-\mathbf{a}_1\|}{2\delta^2}\right), \cdots, \exp\left(\frac{-\|\mathbf{x}-\mathbf{a}_m\|}{2\delta^2}\right)\right]^\top$;

10: Compute the hash function $\mathbf{W}^{(k)}$ by $W^{(k)} = \left(\mathbf{BB}^T + \lambda\mathbf{I}\right)^{-1} 2r\mathbf{BG}^T\mathbf{G}(\mathbf{X}^{(k)})^T \cdots$
$\cdots - r\mathbf{BI}_n^T\mathbf{I}_n(\mathbf{X}^{(k)})^T + \lambda\mathbf{B}(\mathbf{X}^{(k)})^T \cdot \left(\mathbf{X}^{(k)}(\mathbf{X}^{(k)})^T\right)^{-1}$;

11: **return** Hash functions $\{\mathbf{W}^{(k)}\}_{k=1}^{v}$.

2.2 Update V by MGBO

Although Algorithm 2 has a compact and intuitive framework, it could not perform the decent result we expected. Therefore, we adopt an asymmetric strategy by introducing a continuous latent feature matrix $\mathbf{V} \in \mathbb{R}^{r\times n}$ [1]and reformulate the objective function as follows:

$$\begin{aligned} \min_{\mathbf{P},\mathbf{B},\mathbf{Z},\mathbf{V}} \quad & \|\mathbf{PV} - \mathbf{ZL}\|_F^2 + \|\mathbf{L} - \mathbf{ZL}\|_F^2 + \alpha\|\mathbf{V}^\top\mathbf{B} - r\mathbf{S}\|_F^2 + \beta\|\mathbf{B} - \mathbf{V}\|_F^2, \\ \text{s.t.} \quad & \mathbf{B} \in \{-1,1\}^{r\times n},\ \mathbf{VV}^\top = n\mathbf{I}_r,\ \mathbf{V1}_n = \mathbf{0}_r. \end{aligned} \tag{3}$$

The real-valued latent feature $\mathbf{V}$ may preserve more discriminative information by exploring the label information, and it in return guides the hash codes learning via the asymmetric inner product. Besides, the discrete constraint and the binary codes can be generated by simple $sgn(\cdot)$ function.

This time, we employ MGBO to update $\mathbf{V}$, allowing us to update $\mathbf{P}$ and $\mathbf{Z}$ using a continuous matrix instead of a discrete binary matrix. Furthermore, MGBO incorporates global exact penalties[2], which implies that $\mathbf{V}$ can theoretically converge to precise hash codes while perfectly satisfying the orthogonal and equilibrium constraints. Compared with Algorithm 2, Algorithm 3, which utilizes the asymmetric strategy, achieves better performance.

Algorithm 3 (Improved ALECH by MGBO — Update $\mathbf{V}$ by MGBO)

Input: Training data $\{\mathbf{X}^{(k)}\}_{k=1}^{v}$, label matrix $\mathbf{L}$, code length r, model parameters α, β, λ, and maximum iteration number τ.

Output: Hash codes $\mathbf{B}$ and hash functions $\{\mathbf{W}^{(k)}\}_{k=1}^{v}$.

Initialization: Randomly initialize $\mathbf{B}, \mathbf{V}$ with a standard normal distribution, $\mathbf{Z} = \mathbf{I}$;

Compute $\mathbf{G}$ by $\mathbf{G} = \left[\frac{\mathbf{L}_{*1}}{\|\mathbf{L}_{*1}\|_2}, \frac{\mathbf{L}_{*2}}{\|\mathbf{L}_{*2}\|_2}, \ldots, \frac{\mathbf{L}_{*n}}{\|\mathbf{L}_{*n}\|_2}\right]$;

for $i = 1$ to τ **do**

  Update $\mathbf{P}$ by $\mathbf{P} = \mathbf{ZLV}^T/n$;

  Update $\mathbf{Z}$ by $\mathbf{Z} = (\mathbf{PVL}^T + \mathbf{LL}^T)(2\mathbf{LL}^T)^{-1}$;

  Update $\mathbf{V}$ by MGBO;

  Update $\mathbf{B}$ by $\mathbf{B} = sgn(2\alpha r\mathbf{VG}^T\mathbf{G} - \alpha r\mathbf{V}\mathbf{1}_n^T\mathbf{1}_n + \beta\mathbf{V})$;

end for

return Hash codes $\mathbf{B}$.

Map the training data to kernel space by $\phi(\mathbf{x}) = \left[\exp\left(\frac{-\|\mathbf{x}-\mathbf{a}_1\|}{2\delta^2}\right), \cdots, \exp\left(\frac{-\|\mathbf{x}-\mathbf{a}_m\|}{2\delta^2}\right)\right]^\top$;

Compute the hash function $\mathbf{W}^{(k)}$ by $W^{(k)} = \left(\mathbf{BB}^T + \lambda\mathbf{I}\right)^{-1} 2r\mathbf{BG}^T\mathbf{G}(\mathbf{X}^{(k)})^T \cdots$
$\cdots - r\mathbf{BI}_n^T\mathbf{I}_n(\mathbf{X}^{(k)})^T + \lambda\mathbf{B}(\mathbf{X}^{(k)})^T \cdot \left(\mathbf{X}^{(k)}(\mathbf{X}^{(k)})^T\right)^{-1}$;

return Hash functions $\{\mathbf{W}^{(k)}\}_{k=1}^{v}$.

References

[1] Huaxiong Li et al. “Adaptive label correlation based asymmetric discrete hashing for cross-modal retrieval”. In: *IEEE Transactions on Knowledge and Data Engineering* 35.2 (2021), pp. 1185–1199.

[2] Lianghai Xiao, Yitian Qian, and Shaohua Pan. “A relaxation method for binary optimizations on constrained Stiefel manifold”. In: *arXiv preprint arXiv:2308.10506* (2023).