Binary Hashing with Semidefinite Relaxation and Augmented Lagrangian

Thanh-Toan Do, Anh-Dzung Doan, Duc-Thanh Nguyen, Ngai-Man Cheung

Singapore University of Technology and Design
Overview and Contribution

- Two-step approach for hashing
  - inference binary codes
  - learn hash function, given inferred binary codes
    → reduce the complexity; flexible using of different hash functions
- Contribution
  - Unified formulation for supervised/unsupervised hashing
  - Two approaches for inferencing binary codes
    - Semidefinite Programming
    - Augmented Lagrangian
Unified formulation for supervised/unsupervised hashing

- **Input:** $S$: similarity matrix between samples, i.e., pairwise distance/pairwise label matrix for unsupervised/supervised; $L$: code length; $n$: number of training samples
- **Target:** learning binary codes $Z$ s.t. $S$ is preserved in Hamming space, i.e., solving

$$
\min_{Z \in \{-1,1\}^{L \times n}} \|Z^T Z - Y\|^2
$$

- unsupervised: $Y = L - \frac{LS}{2}$; supervised: $Y = L - 2LS$
- Using coordinate descent approach for solving above NP-hard, i.e., solving one row of $Z$ at a time.

Let $x = [x_1, \ldots, x_n]^T = z^{(k)}$, we solve BQP

$$
\min_x x^T Ax
$$

s.t. $x_i^2 = 1, \forall i = 1, \ldots, n.$

(2)

where $A = \{a_{ij}\} \in \mathbb{R}^{n \times n}; a_{ij} = \bar{z}_i^T \bar{z}_j - y_{ij}$. 
Semidefinite Relaxation (SDR) approach

Original BQP: \[
\min_x x^T Ax
\]
\[
s.t. \ x_i^2 = 1, \forall i = 1, ..., n.
\]
Let \( B = A - \lambda_1 I \), where \( \lambda_1 \) is the largest eigenvalue of \( A \); \( X = xx^T \)
→ solve equivalent problem
\[
\min_x \ \text{trace}(BX)
\]
\[
s.t. \ \text{diag}(X) = 1; X \succeq 0; \text{rank}(X) = 1
\] (3)

Two steps solution:
- Drop rank one constraint → solving Semidefinite Program
\[
\min_x \ \text{trace}(BX)
\]
\[
s.t. \ \text{diag}(X) = 1; X \succeq 0
\] (4)

Solving: Using Convex OPT packages: SeDuMi, SDPT3 → achieving the global optimal solution \( X^* \)
- Recover binary solution \( x \) from \( X^* \)
Solving: using randomized rounding process
  - generate \( \xi \) by \( \xi \sim \mathcal{N}(0, X^*) \)
  - get feasible point: \( \hat{x} = \text{sgn}(\xi) \)
Semidefinite Relaxation (SDR) approach (cont’d)

Problem need to be solved:

\[
\min_X \text{trace}(BX)
\]

\[\text{s.t. } \text{diag}(X) = 1; X \succeq 0; \text{rank}(X) = 1 \quad (5)\]

Solution provided SDR-rounding: \(\hat{x}\)

**Bound on objective value at \(\hat{x}\):** Let \(f_{opt}\) be global optimum objective value of (5) and \(f_{SDR-round} = \hat{x}^T B\hat{x}\), we have

\[
f_{opt} \leq E[f_{SDR-round}] \leq \frac{2}{\pi} f_{opt} \quad (6)
\]
Augmented Lagrangian (AL) approach

Original problem:

\[ \min_x x^T A x \]
\[ s.t. \ x_i^2 = 1, \forall i = 1, \ldots, n. \]

Let \( \Phi(x) = [(x_1)^2 - 1, \ldots, (x_n)^2 - 1]^T; \ \Lambda = [\lambda_1, \ldots, \lambda_n]^T \): Lagrange multipliers. Minimizing the unconstrained augmented Lagrangian function

\[ \mathcal{L}(x, \Lambda; \mu) = x^T A x - \Lambda^T \Phi(x) + \frac{\mu}{2} \| \Phi(x) \|^2 \] \quad (7)

- When \( \mu \) is large \( \rightarrow \) penalize the binary constraint violation severely \( \rightarrow \) force the minimizer of the AL function (7) closer to the feasible region of the original problem
- Theoretically, not necessary to take \( \mu \to \infty \) in order to achieve a local optimum of original problem
- CIFAR10: 60k images, 10 categories; GIST features.
- MNIST: 70k images, 10 categories; raw (intensity) features.
- SUN397: 108K images. We only use 42 most frequent categories; AlexNet features.
Evaluation of Supervised/Unsupervised Hashing

Supervised hashing

- **CIFAR10**
- **MNIST**
- **SUN397**

Unsupervised hashing

- **CIFAR10**
- **MNIST**
- **SUN397**
Thank you!