

基于映射字典学习的跨模态哈希检索

姚涛^{1,2} 孔祥维³ 付海燕¹ TIAN Qi⁴

摘要 针对网络上出现越来越多的多模态数据,如何在海量数据中检索不同模态的数据成为一个新的挑战. 哈希方法把数据映射到 Hamming 空间,大大降低了计算复杂度,为海量数据的跨模态检索提供了一条有效的路径. 然而,大部分现存方法生成的哈希码不包含任何语义信息,从而导致算法性能的下降. 为了解决这个问题,本文提出一种基于映射字典学习的跨模态哈希检索算法. 首先,利用映射字典学习一个共享语义子空间,在子空间保持数据模态间的相似性. 然后,提出一种高效的迭代优化算法得到哈希函数,但是可以证明问题的解并不是唯一的. 因此,本文提出通过学习一个正交旋转矩阵最小化量化误差,得到性能更好的哈希函数. 最后,在两个公开数据集上的实验结果说明了该算法优于其他现存方法.

关键词 跨模态检索, 哈希, 映射字典学习, 汉明空间

引用格式 姚涛, 孔祥维, 付海燕, Tian Qi. 基于映射字典学习的跨模态哈希检索. 自动化学报, 2018, 44(8): 1475–1485

DOI 10.16383/j.aas.2017.c160433

Projective Dictionary Learning Hashing for Cross-modal Retrieval

YAO Tao^{1,2} KONG Xiang-Wei³ Fu Hai-Yan¹ TIAN Qi⁴

Abstract With the sharp increasing of multi-modal data on the internet, retrieving samples with different modalities has become a challenge. hashing methods, which map the data to Hamming space to reduce computational cost, can provide an effective way for large-scale cross-modal retrieval. However, the hashing codes of most existing methods do not contain any semantic information, which degrades the performance. To address this issue, we propose a cross-modal hashing method, termed projective dictionary learning hashing (PDLH) method. Firstly, projective dictionary learning is employed to learn a sharing semantic subspace by preserving the inter-modal similarity. Then an efficient iterative optimal algorithm is proposed to gain hashing functions. However, the solution is not a unique solution, as proven in this paper. To further improve the performance of the proposed method, an orthogonal rotation matrix is learned by minimizing the quantization loss for better hashing functions. Finally, experimental results on two widely used datasets show that the performance of the proposed method is better than those of the existing methods.

Key words Cross-modal retrieval, hashing, project dictionary learning, Hamming space

Citation Yao Tao, Kong Xiang-Wei, Fu Hai-Yan, Tian Qi. Projective dictionary learning hashing for cross-modal retrieval. *Acta Automatica Sinica*, 2018, 44(8): 1475–1485

随着计算机网络和信息技术的快速发展,网络

上的媒体数据量急剧增长,媒体的表示形式也呈现出多模态性(图像、文本、声音、视频等). 例如:在微博上传照片时,往往会同时上传一段文字描述照片的内容或用一些标签标注图像的内容;在微信的朋友圈分享时,往往也是图文并茂;购物网站,例如淘宝、京东等,在描述产品信息时通常既用图片,又用文字. 这些多模态数据虽然表现形式不同,但它们之间存在语义关联. 跨媒体检索的目的就是挖掘不同媒体之间存在的语义关系,并按语义关系进行排序,返回跟查询存在较强语义关系的不同模态的数据. 随着媒体数据量的急速增长和模态的多样化,传统的检索方法已经不能满足当前跨媒体检索的需求. 如何在海量数据中检索不同模态的数据成为一个巨大的挑战.

哈希方法是解决大数据问题的一种有效的方法,不仅能大大节省存储空间,而且计算效率也大幅提高. 例如一张图片用 5 000 维的 BOW 表示,假设每

收稿日期 2016-05-27 录用日期 2017-04-21

Manuscript received May 27, 2016; accepted April 21, 2017

国家自然科学基金(71421001, 61502073, 61172109, 61429201), 模式识别国家重点实验室开放课题(201407349)资助

Supported by National Natural Science Foundation of China (71421001, 61502073, 61172109, 61429201) and the Open Projects Program of National Laboratory of Pattern Recognition (201407349)

本文责任编辑 朱军

Recommended by Associate Editor ZHU Jun

1. 大连理工大学信息与通信工程学院 大连 116024 中国 2. 鲁东大学信息与电气工程学院 烟台 264025 中国 3. 浙江大学数据科学与管理工程学系 杭州 310058 中国 4. 德克萨斯大学圣安东尼奥分校计算机科学学院 圣安东尼奥 78249 美国

1. Department of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China

2. Department of Information and Electrical Engineering, Ludong University, Yantai 264025, China

3. Department of Data Science and Engineering Management, Zhejiang University, Hangzhou 310058, China

4. Department of Computer Science, University of Texas at San Antonio, San Antonio 78249, USA

维用 double 数据类型表示, 即每维占用 8 Bytes 的存储空间, 则需要 $5000 \times 8 \text{ Bytes} = 40000 \text{ Bytes}$ 的存储空间. 而哈希方法是把数据映射汉明空间, 例如用 32 bits (8 bits = 1 Byte) 来表示一张图片, 仅需要 4 Bytes 的存储空间, 大大节省了存储空间, 占用的存储空间仅为原始特征空间的万分之一. 在检索过程中, 因为数据用二值码表示, 因此在计算样本间的距离时, 只需要做简单的异或操作即可, 大大提升了计算的效率, 使检索的时间复杂度远低于传统方法.

针对以上问题, 本文提出一种基于映射字典学习的跨媒体哈希算法, 主要贡献如下:

1) 利用映射字典学习使哈希码含有语义信息以提升算法的性能. 算法同时学习了哈希函数, 这与现存的字典学习哈希算法不同.

2) 提出通过最小化量化误差, 学习一个正交旋转矩阵, 提升算法的性能, 并且可以证明旋转后的解依然是问题的局部最优解.

本文结构安排如下: 第 1 节介绍哈希算法的相关工作; 第 2 节回顾字典学习的相关内容, 阐述了本文算法的思想, 优化过程及计算复杂度分析; 第 3 节给出在两个公开数据集上的实验结果; 第 4 节对本文的研究内容进行总结.

1 相关工作

由于哈希方法的高效性和节省内存, 最近引起了越来越多的关注^[1-15]. 哈希方法一般可以分为单模态哈希和多模态哈希.

单模态哈希方法又可以分为不依赖数据的哈希方法和依赖数据的哈希方法. 在不依赖数据的哈希方法中, 最先提出的是局部敏感哈希 (Local sensitive hashing, LSH), 它利用随机线性映射作为哈希函数, 把原始空间中的样本映射到汉明空间^[1]. 但是, 样本之间的相似性是非线性的, 线性函数很难捕捉样本之间的内在联系, 因此核局部敏感哈希 (Kernelized local sensitive hashing, KLSH) 提出利用核方法捕捉样本之间的内在联系^[2]. 但是, 这类算法往往需要很长的哈希码和多个哈希表才能取得较好的实验结果. 然而随着哈希码长度的增加, 会降低相似样本映射到同一个桶的概率, 导致召回率的迅速降低, 而且较长的哈希码必然会增加存储空间和计算复杂度. 相对于不依赖于数据的哈希方法, 依赖于数据的哈希方法可以获得更为紧凑的表示, 较小的码长就可以获得令人满意的结果, 因此受到越来越多的研究人员关注^[4-8]. 谱哈希 (Spectral hashing, SH) 通过放松哈希函数的二值约束, 利用谱图分析和拉普拉斯特征函数学习哈希函数^[4]. 核监督哈希

(Supervised hashing with kernels, KSH) 利用核方法学习哈希函数, 在汉明空间最小化正样本对之间的距离, 最大化负样本对的距离^[5]. 以上算法取得较好的实验结果, 但是没有考虑量化损失的影响, 导致学习到的哈希函数的性能的下降. 迭代量化哈希 (Iterative quantization, ITQ) 通过最小化量化误差, 学习一个旋转矩阵, 得到性能更好的哈希函数^[6]. 监督离散哈希 (Supervised discrete hashing, SDH) 提出了一种离散优化算法, 直接可以得到问题的离散局部最优解^[7]. 然而随着网络上多模态数据的不断增长, 一个网页可以包含多种模态的数据, 而单模态的哈希方法通常不能直接用于多模态数据, 如何把多模态数据纳入到一个统一的学习框架成为新的挑战.

多模态哈希方法一般分为多模态融合哈希方法和跨模态哈希方法, 本文主要研究跨模态哈希方法. 多模态融合哈希方法利用不同特征之间的互补性, 学习一个更好的汉明空间, 提升算法的性能^[8-10]. 跨模态哈希的目标是学习一个共享的汉明空间, 在这个空间可以实现跨媒体检索^[6, 11-17]. 基于相似敏感哈希的跨模态度量学习方法 (Cross-modality metric learning using similarity sensitive hashing, CMSSH) 通过最小化不同模态的相似样本之间的汉明距离, 最大化不同模态的不相似样本间的汉明距离, 学习哈希函数^[11]. 典型相关分析 (Canonical correlation analysis, CCA)^[18] 哈希方法, 把 CCA 引入跨媒体哈希方法, 提出最大化模态间的相关性, 学习一组哈希函数^[6]. 然而上述方法只保持了模态间的相似性, 忽视了模态内的相似性. 跨视角哈希 (Cross view hashing, CVH) 把谱哈希扩展到跨模态检索, 通过最小化加权距离, 保持相似样本 (模态内和模态间) 的相似性^[12]. 多模态潜在二值嵌入 (Multi-modal latent binary embedding, MLBE) 提出一个概率生成模型, 通过保持多模态样本的模态内和模态间的相似度来学习哈希函数^[16]. 然而, 这些方法并没有为不同模态的数据学习一个统一特征空间, 导致无法捕捉不同模态间存在的一些潜在的共享信息. 协同矩阵分解哈希 (Collective matrix factorization hashing, CMFH) 利用协同矩阵分解保持模态间的相似性, 为样本对学习同一表示^[13]. 基于聚类联合矩阵分解哈希 (Cluster-based joint matrix factorization hashing, CJMFH) 提出了首先对各个模态进行聚类运算, 再利用矩阵分解同时保持模态内、模态间和基于聚类的相似性^[17]. 以上方法虽然取得了令人满意的结果, 但是学习到的哈希码不包含任何语义信息, 限制了算法的性能. 稀疏哈希 (Latent semantic sparse hashing, LSSH) 为

了缩小图像和文本之间的语义鸿沟, 利用稀疏表示学习图像的一些显著结构, 利用矩阵分解为文本学习一个潜在的语义空间, 并保持模态间的语义相似性^[14]. 稀疏多模态哈希 (Sparse multi-modal hashing, SMMH) 提出利用稀疏表示为图像和文本学习一个共同的语义空间, 保持模态间的相似性^[15]. 这类方法利用稀疏表示, 使哈希码包含语义信息, 提升了算法的性能. 但是这类算法通常存在以下问题, 限制了算法的应用. 1) 在字典学习算法中, 因为稀疏约束项的存在, 导致训练和测试过程算法的复杂度. 2) 这些哈希算法不同于传统的哈希算法, 没有学习显式的哈希函数. 测试样本, 通常需要首先解决一个 Lasso 问题, 得到样本的稀疏表示, 然后通过量化得到样本的哈希码 (例如文献 [14]), 而不能像传统算法那样直接利用哈希函数到. 3) 因为得到的系数矩阵是稀疏的, 导致了哈希码的 0 和 1 分配不均匀.

针对以上问题, 本文提出一种基于映射字典学习的哈希算法. 在字典学习过程中, 放松了稀疏约束项算法, 不仅降低了时间复杂度和平衡了哈希码的分布, 而且在字典学习过程中得到了哈希函数. 对于哈希问题的求解, 现存的大部分跨模态哈希算法往往采用谱松弛的方法得到连续的最优解^[11, 19], 没有考虑量化损失对算法性能的影响, 而导致性能的下降^[3]. 本文受 ITQ 的启发, 通过最小化量化误差, 学习一个正交的旋转矩阵, 进一步提升算法的性能.

2 映射字典学习哈希算法 (DPLH)

本节首先介绍字典学习的基本内容和本文提出的基于映射字典学习的哈希算法, 然后引出优化算法及正交变换的相关内容, 最后分析算法的时间复杂度, 证明算法的高效性.

2.1 映射字典学习

字典学习已经在图像处理和计算机视觉领域取得了巨大的成功^[20-28]. 传统的字典学习可以分为综合字典学习和分析字典学习两类. 综合字典学习应用在各个领域取得了令人满意的成果^[21-27], 而分析字典学习的研究正处在起步阶段^[28].

综合字典学习的目标函数一般定义为

$$\min_{D, A} \|X - DA\|_F^2 + \lambda \|A\|_p \quad (1)$$

其中, $X \in \mathbf{R}^{d \times N}$ 表示数据, $D \in \mathbf{R}^{d \times c}$ 表示字典 (字典的每行 $D(:, i)$ 称为字典的一个原子), $A \in \mathbf{R}^{c \times N}$ 表示系数矩阵, λ 为权重参数, $\|\cdot\|_F$ 表示 Frobenius 范数, 一般情况下取 $p = 1$, 即 l_1 范数. 式 (1) 表明数据 X 可以由字典和稀疏的系数矩阵重构.

分析字典学习目标函数一般定义为

$$A = \min_{\Omega} \|\Omega X\|_p \quad (2)$$

这里的范数可以为 l_1 或 l_2 范数, A 为稀疏矩阵, Ω 表示字典.

然而, 在字典学习过程中加入 l_1 或 l_2 稀疏约束项, 往往会导致过大的计算量. 文献 [27] 把综合字典学习和分析字典学习纳入同一个学习框架, 提出了一种基于映射字典学习的分类方法, 利用线性映射代替非线性的稀疏编码, 取得了令人满意的结果. 受此启发, 本文利用线性映射的方法来进行字典学习, 以减少时间复杂度, 同时把学习的线性映射作为哈希函数. 映射字典学习的目标函数定义为

$$\begin{aligned} \min_{D, P} \|X - DPX\|_F^2 + \text{Re}(P) \\ \text{s.t.} \quad \|D(:, i)\|_F^2 \leq 1 \end{aligned} \quad (3)$$

其中, $\text{Re}(\cdot)$ 表示正则项, P 可以看作重构矩阵.

2.2 DPLH 算法的思想

在描述 DPLH 算法之前, 先对本文用到的符号进行说明. 为了描述方便, 本文只考虑两种模态, 例如: 图像和文本, 当然算法可以很容易扩展到多于两种模态的情况. 用 $X^{(k)} = \{x_1^{(k)}, x_2^{(k)}, \dots, x_N^{(k)}\}$ 表示第 k 个模态的特征描述, $k = 1, 2$. $X^{(k)} \in \mathbf{R}^{d_k \times N}$, $\{x_i^{(1)}, x_i^{(2)}\}$ 表示第 i 个样本由不同模态描述构成的样本对. 其中, d_k 表示第 k 个模态的特征空间的维数, N 表示样本对的数量. 用 $A_k \in \mathbf{R}^{c \times N}$ 表示第 K 个模态的系数矩阵, $D_k \in \mathbf{R}^{d_k \times c}$ 表示第 k 个模态的字典, $P_k \in \mathbf{R}^{c \times d_k}$ 表示第 k 个模态的哈希函数 (即上面提到的重构矩阵), $B^{(k)} \in \{0, 1\}^{c \times N}$ 表示第 k 个模态的哈希码, 其中 c 表示哈希码的长度. 本文把两个模态纳入到一个学习框架中, 则映射字典学习算法的目标函数定义为

$$\begin{aligned} \min_{D_1, D_2, P_1, P_2} (1 - \lambda) \|X^{(1)} - D_1 P_1 X^{(1)}\|_F^2 + \\ \lambda \|X^{(2)} - D_2 P_2 X^{(2)}\|_F^2 \\ \text{s.t.} \quad \|D_1(:, i)\|_F^2 \leq 1, \|D_2(:, i)\|_F^2 \leq 1 \end{aligned} \quad (4)$$

其中, 前两项是重构误差, β, λ 为权重参数, $D_1(:, i)$ 表示字典 D_1 的第 i 个字典原子, $D_2(:, i)$ 表示字典 D_2 的第 i 个字典原子.

跨媒体检索的目标是学习一个低维的共享子空间, 异构数据之间的相似度可以在此空间直接度量. 样本对虽然用不同模态表示, 但它们包含相同的语义信息, 因此在学习的子空间中, 它们的差异应该尽量小. 文献 [13] 把协同矩阵分解引入到子空间学习,

但是样本对在子空间中相同表示的强约束条件, 可能会导致算法性能的下降. 因此, 本文放松了此约束, 目标函数定义为

$$\begin{aligned} \min_{D_1, D_2, P_1, P_2} & (1 - \lambda) \|X^{(1)} - D_1 P_1 X^{(1)}\|_F^2 + \\ & \lambda \|X^{(2)} - D_2 P_2 X^{(2)}\|_F^2 + \\ & \mu \|P_1 X^{(1)} - P_2 X^{(2)}\|_F^2 + \\ & \beta \text{Re}(P_1, P_2) \\ \text{s.t.} & \|D_1(:, i)\|_F^2 \leq 1, \|D_2(:, i)\|_F^2 \leq 1 \end{aligned} \quad (5)$$

2.3 迭代优化算法

为了更容易求解式 (4), 为两个模态分别引入一个中间变量 A_1 和 A_2 , 目标函数可写为

$$\begin{aligned} \min_{D_1, D_2, P_1, P_2, A_1, A_2} & (1 - \lambda) \|X^{(1)} - D_1 A_1\|_F^2 + \\ & \lambda \|X^{(2)} - D_2 A_2\|_F^2 + \mu \|A_1 - A_2\|_F^2 + \\ & \alpha \left(\|A_1 - P_1 X^{(1)}\|_F^2 + \|A_2 - P_2 X^{(2)}\|_F^2 \right) + \\ & \beta \text{Re}(P_1, P_2) \\ \text{s.t.} & \|D_1(:, i)\|_F^2 \leq 1, \|D_2(:, i)\|_F^2 \leq 1 \end{aligned} \quad (6)$$

其中, 参数 α 为权重.

式 (6) 的求解是一个非凸优化问题. 幸运的是, 当求解一个变量而固定其他变量时, 问题就变为凸的, 所以可以利用迭代的方法求解.

1) 固定其他变量求解 A_1 , 则式 (6) 可写为

$$\begin{aligned} \ell(A_1) = & (1 - \lambda) \|X^{(1)} - D_1 A_1\|_F^2 + \\ & \alpha \|A_1 - P_1 X^{(1)}\|_F^2 + \\ & \mu \|A_1 - A_2\|_F^2 + \text{const} \end{aligned} \quad (7)$$

令 $\frac{d\ell(A_1)}{dA_1} = 0$, 则可以得到闭合解.

$$\begin{aligned} A_1 = & \left(D_1^T D_1 + \frac{(\alpha + \mu)}{(1 - \lambda)} I \right)^{-1} \times \\ & \left(\frac{\alpha}{(1 - \lambda)} P_1 X^{(1)} + \frac{\mu}{(1 - \lambda)} A_2 + D_1^T X^{(1)} \right) \end{aligned} \quad (8)$$

同理

$$\begin{aligned} A_2 = & \left(D_2^T D_2 + \frac{(\alpha + \mu)}{\lambda} I \right)^{-1} \times \\ & \left(\frac{\alpha}{\lambda} P_2 X^{(2)} + \frac{\mu}{\lambda} A_1 + D_2^T X^{(2)} \right) \end{aligned} \quad (9)$$

2) 固定其他变量求解 P_1 , 则式 (5) 可写为

$$\min_{P_1} \|A_1 - P_1 X^{(1)}\|_F^2 + \frac{\beta}{\alpha} \|P_1\|_F^2 + \text{const} \quad (10)$$

展开上式并对 P_1 求导, 令其导数为零, 可以得到闭合解.

$$P_1 = A_1 X^{(1)T} \left(X^{(1)} X^{(1)T} + \frac{\beta}{\alpha} I \right)^{-1} \quad (11)$$

同理

$$P_2 = A_2 X^{(2)T} \left(X^{(2)} X^{(2)T} + \frac{\beta}{\alpha} I \right)^{-1} \quad (12)$$

3) 固定其他变量求解 D_1 , 则式 (5) 可写为

$$\begin{aligned} \min_{D_1} & \|X^{(1)} - D_1 A_1\|_F^2 + \text{const} \\ \text{s.t.} & \|D_1(:, i)\|_F^2 \leq 1 \end{aligned} \quad (13)$$

式 (13) 可以用文献 [27] 提出的 ADMM 算法求解, 同理 D_2 也可以用相同方法求解.

上述过程不断迭代, 直到目标函数收敛为止.

2.4 学习正交旋转矩阵

在得到哈希函数 P_1, P_2 后, 测试样本的哈希码可以通过哈希函数直接得到.

$$b_j^{(k)} = \frac{\text{sgn} \left(P_k x_j^{(k)} - \overline{P_k X^{(k)}} \right) + 1}{2} \quad (14)$$

其中, $\text{sgn}(\cdot)$ 表示符号函数, $\overline{P_k X^{(k)}}$ 表示第 k 个模态映射到子空间的样本均值. 在这里减去均值是为了保证哈希码 0 和 1 分布均匀.

式 (14) 表示对于 P_k 的任意一行 $P_k(i)$, 如果满足 $P_k(i, :) x_j^{(k)} - \overline{P_k(i, :) X^{(k)}} > 0$, 则 $b_j^{(k)}(i) = 1$, 否则 $b_j^{(k)}(i) = 0$. 然而式 (14) 的量化运算会带来量化损失, 而损失的大小会直接影响算法的性能, 通常量化损失越小越好. 但是, 大部分现存的哈希算法, 直接利用式 (14) 得到哈希码, 而没有考虑量化损失对哈希函数性能的影响^[1-5, 13-14, 19]. 文献 [6] 提出了通过最小化量化误差学习一个旋转矩阵, 以得到性能更好的哈希函数, 提升了算法的性能. 受此启发, 本文得到哈希函数 P_1, P_2 后, 通过最小化量化误差, 学习一个正交变换矩阵, 得到性能更好的哈希函数. 量化产生的损失定义为

$$\sum_{i=1}^2 \|B^{(i)} - R V_i\|_F^2 \quad \text{s.t.} \quad R^T R = I_c \quad (15)$$

其中, $V_i = P_i X^{(i)} - \overline{P_i X^{(i)}}$, $B^{(i)} = \text{sgn}(V_i)$. 式 (15) 可以利用迭代的方法求解.

固定 R , 求 $B^{(i)}$.

$$B^{(i)} = \text{sgn}(R V_i) \quad (16)$$

固定 $B^{(i)}$, 求 R .

$$\min_{R^T R = I_c} \sum_{i=1}^2 \|B^{(i)} - RV_i\|_F^2 \quad (17)$$

式 (17) 是典型的 Orthogonal Procrustes problem, 可以由奇异值分解 (Singular value decomposition, SVD) 的方法解决. 为了最小化量化误差, 哈希函数更新为 RP_i .

下面证明 RP_i 不仅可以最小化量化误差, 而且同时是目标函数 (式 (5)) 的局部最优解, 即正交不变定理.

定理 1. 设 R 是 $c \times c$ 的可逆正交变换矩阵, 满足 $R^T R = I_c$. 如果 $P_1, P_2, D_1, D_2, A_1, A_2$ 是式 (6) 的局部最优解, 则 $RP_1, RP_2, D_1 R^T, D_2 R^T, RA_1, RA_2$ 也是式 (6) 的优化解.

证明.

$$\begin{aligned} \|X^{(1)} - D_1 R^T R A_1\|_F^2 &= \|X^{(1)} - D_1 A_1\|_F^2 \\ \|X^{(2)} - D_2 R^T R A_2\|_F^2 &= \|X^{(2)} - D_2 A_2\|_F^2 \\ \|RA_1 - RP_1 X^{(1)}\|_F^2 &= A_1^T R^T R A_1 - \\ &\quad A_1^T R^T R P_1 X^{(1)} - X^{(1)T} P_1^T R^T R A_1 + \\ &\quad X^{(1)T} P_1^T R^T R P_1 X^{(1)} = \|A_1 - P_1 X^{(1)}\|_F^2 \end{aligned}$$

同理,

$$\begin{aligned} \|RA_2 - RP_2 X^{(2)}\|_F^2 &= \|A_2 - P_2 X^{(2)}\|_F^2 \\ \|RA_1 - RA_2\|_F^2 &= \|A_1 - A_2\|_F^2 \end{aligned}$$

□

定理 1 证明了式 (6) 的局部最优解并不是唯一的, 存在正交变换矩阵 R , 使 RP_i 也是式 (6) 的一个局部最优解, 因此直接优化式 (6) 得到的解并不一定是问题的最优解. 本文通过最小化量化损失学习一个正交变换矩阵 R , 使得 RP_i 既是式 (6) 的局部最优解, 又满足量化损失最小, 提升了算法的性能.

2.5 复杂度分析

在训练过程中, 计算复杂度包括两部分: 目标函数的求解和正交旋转矩阵的求解.

目标函数的求解过程是迭代优化的过程, 不断迭代更新 $P_1, P_2, D_1, D_2, A_1, A_2$, 直到算法收敛, 因此训练的计算复杂度主要产生在迭代更新过程. 在这里, 首先分析一下更新每个变量的计算复杂度. 在更新变量 $A_i, i = 1, 2$ 的表达式中, 第 1 项计算复杂度为 $O(c^2 d_i + c^3)$, 第 2 项的计算复杂度为 $O(cd_i N)$, 因此更新 A_i 的计算复杂度为 $O(c^2 d_i + c^3 + cd_i N + c^2 N)$. 对于变量 P_i , 通过观察发现, P_i

包含常数项 $(X^{(i)} X^{(i)T} + \beta I)^{-1}$, 计算它们的时间复杂度为 $O(d_i^2 N + d_i^3)$. 但只需要在迭代前计算一次, 并存储, 在迭代过程中, 只需读取即可, 从而降低计算复杂度. 因此迭代 P_i 的计算复杂度为 $O(cd_i N + cd_i^2)$. 利用 ADMM 算法更新 D_i 的计算复杂度为 $O(d_i N + c^3 + c^2 d_i + cd_i^2)$.

正交旋转矩阵的求解也是利用迭代算法. 其中, 更新 R 首先要计算 $B^{(i)} V^T$ 的 SVD 分解, 即 $B^{(i)} V^T = S \Omega \hat{S}^T$, 而 $R = \hat{S}^T S$, 所以总的计算复杂度为 $O(c^2 N + c^3)$. 而更新 $B^{(i)}$ 的时间复杂度为 $O(c^2 N)$.

在大数据时代, 数据量非常大. 通常在实际应用中 N 的值很大, 一般情况下, N 远远大于 d_i 和 c 的值. 因此, 整个迭代训练过程的计算复杂度为 $O(N)$, 即与训练数据集的大小是线性关系. 训练过程的计算复杂度低, 保证了算法的可扩展性.

而对于测试过程, 因为本文生成了哈希函数, 测试样本的哈希码可以直接通过哈希函数得到, 所以两个模态的计算复杂度分别为 $O(cd_1)$ 和 $O(cd_2)$. 而检索过程为求哈希码的距离, 可以通过高效的异或运算实现. 因此, 测试过程的计算复杂度也很低.

3 实验结果

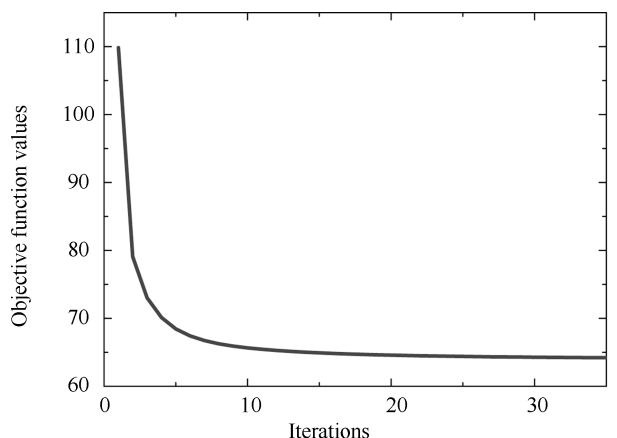
在实验中, 本文主要通过两个检索任务验证算法的有效性: 利用图像检索文本和利用文本检索图像. 通过实验结果发现, 本文提出的无监督算法, 在某些情况下取得了优于监督算法的性能, 证明了算法的有效性.

PDLH 有 4 个参数, 参数 λ 是一个平衡参数, 控制两个模态的权重, 在实验中发现这个参数鲁棒性较强, 本文设置 $\lambda = 0.5$, 表明两个模态的重要性相同. 参数 α 控制重构系数矩阵产生的损失的权重, 此参数也具有一定的鲁棒性, 根据经验本文取 $\alpha = 0.3$. 参数 μ 是模态间相似性保持的权重, 在跨模态检索中 μ 的作用较大, 因此应该取较大的值, 根据经验本文设置 $\mu = 2$. 参数 β 是正则化项的权重, 因此应该取较小的值, 根据经验本文设置 $\beta = 0.02$.

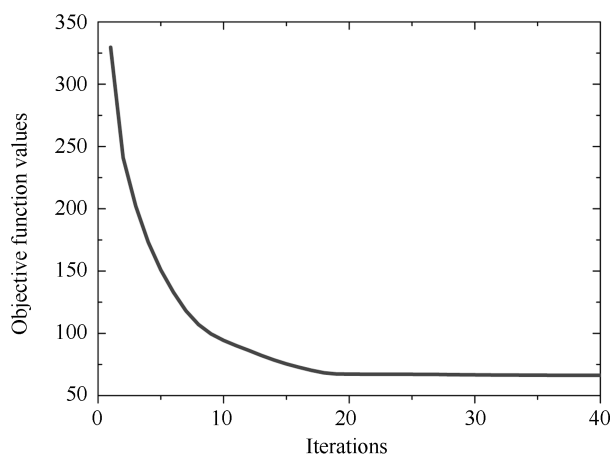
为了验证迭代优化算法的有效性, 本文在 Wiki 和 NUS-WIDE 数据集上进行了实验 (哈希码长为 32 bits), 实验结果如图 1 所示. 通过图 1 发现本文提出的优化算法收敛速度很快, 在少于 20 次迭代后便收敛, 说明了优化算法的有效性.

3.1 实验数据集

本文在 Wiki^[29] 和 NUS-WIDE^[30] 两个公开数据集上进行实验, 并与现存算法比较, 以验证本文算法性能.



(a) 在 Wiki 数据集上的结果
(a) Result on Wiki dataset



(b) 在 NUS-WIDE 数据集上的结果
(b) Result on NUS-WIDE dataset

图1 算法的收敛性分析
Fig. 1 Convergence analysis of the proposed optimization algorithm

Wiki 数据集: 包含 2 866 个文档, 这些文档是从维基百科收集的, 每个文档由一张图片和描述它的一段文本组成, 这些文档可以分为 10 类. 数据集中每张图片用 128 维的 BOW 特征向量表示, 而每段文本用 10 维的 Latent Dirichlet allocation (LDA) 特征向量表示. 其中, 随机选择 75 % 的文档构成训练集, 剩余的 25 % 构成测试集.

NUS-WIDE 数据集: 包含 269 648 张图片, 这些图片是从 Flickr 上收集的. 每张图片与它对应的标注词构成图像-文本对, 每张图片平均有 6 个标注词. 这些样本对被分为 81 个类, 然而有些类的样本数量很少, 为了保证每类有足够多的训练样本, 本文选取了样本数量最多的 10 个类, 186 577 个样本对. 其中图片用 500 维的 BOW 向量表示, 而文本用 1 000 维的 BOW 向量表示. 参照文献 [13, 19] 的

设置, 本文从数据集中随机选择 99 % 的图像-文本对构成训练集, 剩余的 1 % 构成测试集.

3.2 对比算法及评价标准

为了验证 PDLH 算法的有效性, 将 PDLH 与现有算法在两个公开数据集上的实验结果进行对比, 现有算法包括 CCA^[6], CVH^[12], CMFH^[13], SCM^[19], LSSH^[14] 和 STMH^[31]. 其中文献 [19] 使用了两种优化算法: 正交优化算法和序列优化算法, 本文分别用 SCM-O 和 SCM-S 表示. 并且 SCM 算法利用标签建立相似矩阵, 所以为监督的跨模态哈希算法, 而其他方法为无监督算法. 为了验证本文提出的利用映射字典学习子空间方法的有效性, 本文利用 PDLH- 表示去除旋转矩阵的实验结果.

在实验中, 因为 STMH 算法的代码没有公开, 所以由我们实现, 而其余对比算法的代码都由作者提供. 所有代码的参数都经过调试, 并且我们报告的是最好的实验结果. 在实验中, 本文以标签作为判定标准, 即两个样本的标签至少含有一个相同的类, 才判定这两个样本为同一类.

在实验中, 本文利用广泛应用于哈希算法的 Mean average precision (MAP) 来评估各算法的性能. Average precision (AP) 的定义如下:

$$AP = \frac{1}{L} \sum_{i=1}^r prec(i) \times \delta(i) \quad (18)$$

其中, r 为检索到的样本数量, L 为检索到的正确样本的数量, $prec(i)$ 为前 i 个样本的准确率, $\delta(i)$ 为指示函数, 当第 i 个样本为正确样本时 $\delta(i) = 1$, 否则 $\delta(i) = 0$. 而 MAP 为所有测试样本 AP 的平均.

本文用每个测试样本检索返回的前 200 个样本计算 MAP, 记为 MAP@200. 为了进一步证明 PDLH 的有效性, 本文绘制了 Precision-recall (PR) 曲线图, PR 曲线反映了不同正确样本召回率对应的准确率.

3.3 Wiki 数据集的实验结果

由于 Wiki 数据集较小, 所以本文利用所有训练集的样本训练哈希函数. 在实验中, 本文测试了不同哈希码长的算法性能, 其中 MAP 的实验结果见表 1. 从表 1 可以看出, 1) PDLH, CMFH 和 SCM-S 算法的性能较好, PDLH 在大多数码长取得了最好的实验结果, 只在少数码长低于 SCM-S 或 CMFH 的性能; 2) 在所有码长情况下, PDLH 的结果都优于 PDLH- 的结果. 这证明了正交旋转矩阵通过最小量化误差提升了算法的性能; 3) 即使在去除旋转矩阵的情况下, 本文提出的算法也取得了良好的性能. 这说明利用映射字典学习不仅降低了算法的时

表 1 图像检索文本和文本检索图像任务在 Wiki 数据集上的实验结果 (MAP@200)

Table 1 MAP@200 results on Wiki dataset for the tasks of using the image to query texts and vice versa

算法	任务	8	16	24	32	任务	8	16	24	32
CCA		0.2047	0.1815	0.1634	0.1694		0.2036	0.1663	0.1527	0.1595
CVH	图	0.2038	0.1951	0.1682	0.1674	文	0.1997	0.1833	0.1703	0.1613
SCM-O	像	0.1907	0.1718	0.1673	0.1704	本	0.1889	0.1669	0.1610	0.1661
SCM-S	检	0.2129	0.2353	0.2337	0.2377	检	0.2037	0.2411	0.2419	0.2507
CMFH	索	0.2185	0.2300	0.2377	0.2420	索	0.2216	0.2333	0.2352	0.2390
LSSH	文	0.1895	0.2084	0.2232	0.2094	图	0.1841	0.2127	0.2308	0.2157
STMH	本	0.1907	0.1926	0.2201	0.2321	像	0.1896	0.2130	0.2260	0.2240
PDLH-		0.2188	0.2175	0.2385	0.2316		0.2217	0.2162	0.2364	0.2325
PDLH		0.2196	0.2301	0.2499	0.2384		0.2225	0.2276	0.2423	0.2430

间复杂度, 而且通过字典学习, 使哈希码含有语义信息, 增强了哈希码的区分能力, 因此得到了令人满意的实验结果.

SCM-S 是监督哈希算法, 在哈希函数学习过程中不仅利用了特征信息, 而且利用了所有样本的标签信息来获得更好的哈希函数. 然而, 获得所有样本的标签, 要耗费大量的人力物力, 在大数据时代, 是不可能实现的, 所以本文算法更具有应用价值.

为了进一步证明本文提出算法的有效性, 图 2 和图 3 分别绘制了码长为 16 bits 和 32 bits 时各个算法在两个任务上的 PR 曲线图. 从图 2 和图 3 可以看出, PDLH 和 SCM-S 算法取得了较好的性能, 而 PDLH 在召回率较低时性能更好一些. 这在实际应用中非常重要, 因为用户在检索时, 往往更关注排列在前的返回样本.

3.4 NUS-WIDE 数据集的实验结果

由于 NUS-WIDE 的训练集较大, 而 LSSH 和 SCM-O 需要大量的训练时间. 为了降低时间复杂度, 参考文献 [13] 的参数设置, 本文从训练集中随机选出 5 000 个样本对构成训练集, 而测试集包含的 1% 的样本对全部用作测试. MAP 的实验结果见表 2. 从表 2 可以看出, 在图像检索文本任务中, PDLH 在各码长都得到了最好的结果, 而且性能明显优于其他算法. 而在文本检索图像任务中 PDLH 和监督算法 SCM-S 取得了明显优于其他算法的结果. 而且即使在去掉旋转矩阵的情况下, 本文算法依然在大多数情况下取得了最好结果. 实验结果再次证明了算法既降低了复杂度, 又提升了子空间的区分能力. 同时也验证了使哈希码含有语义信息提升了算法的

性能 (在大部分情况下, 性能超过监督算法 SCM-S, 其余情况下, 性能逼近监督算法 SCM-S).

为了进一步证明本文提出算法的有效性, 图 4 和图 5 分别绘制了码长为 16 bits 和 32 bits 时各个算法在两个任务上的 PR 曲线图. 从图 4 和图 5 可以看出, 与 MAP 的结果类似, PDLH 算法和监督算法 SCM-S 的性能在 NUS-WIDE 上明显优于现有的其他的无监督算法. 而与监督算法 SCM-S 相比, PDLH 在召回率较低时性能较好, 这与在 Wiki 数据集上的结果类似.

为了进一步验证本文算法的可扩展性, 本文设定哈希码码长为 16 bits, 并对训练集的大小进行了不同设定, 训练时间和 MAP 的实验结果见表 3 所示. 从表 3 可以看出, 随着训练集样本数量的增加, 算法的性能不断提升. 这是很合理的, 因为随着训练集样本数量的增多, 训练样本包含样本间的内在联系信息越丰富, 因此可以学习更好的哈希函数. 而且通过表 3 还发现训练时间与样本的尺寸基本呈线性关系, 从实验上验证了本文之前的复杂度分析.

4 结论

针对哈希码语义无关而导致性能下降的问题, 本文提出了一种基于映射字典学习的跨模态哈希检索算法. 算法利用映射字典学习降低了算法复杂度, 并生成了哈希函数, 这与现存字典学习哈希方法不同. 最后在两个公开数据集上的实验结果证明了算法的有效性. 将来的工作主要包括学习一个更好的子空间表示, 减小量化误差对哈希函数的影响和利用非线性变换更好地捕捉样本间的内在联系.

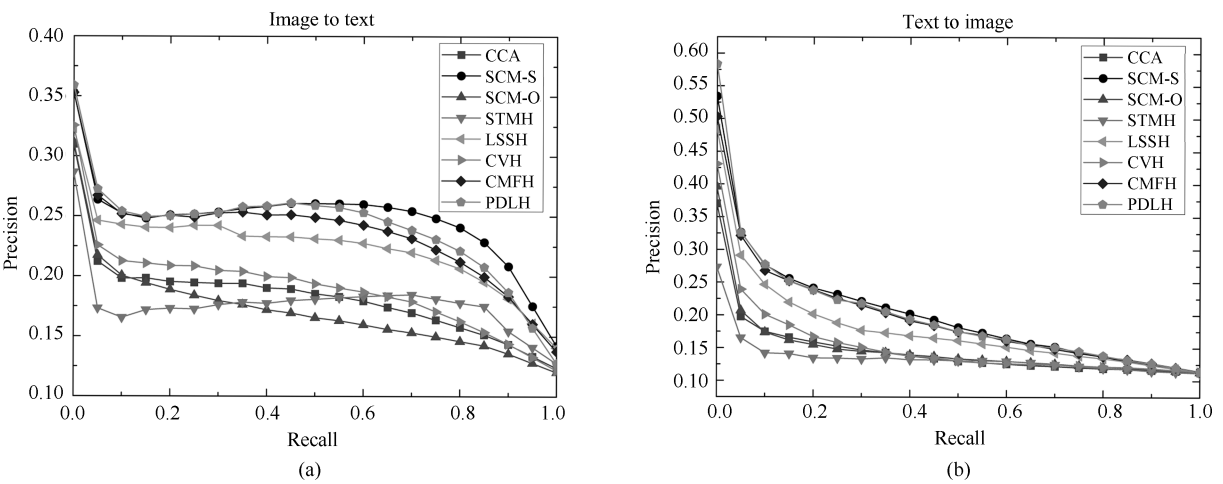


图 2 码长 16 bits 在 Wiki 数据集的 PR 曲线图

Fig. 2 PR curves on Wiki dataset with the code length fixed to 16 bits

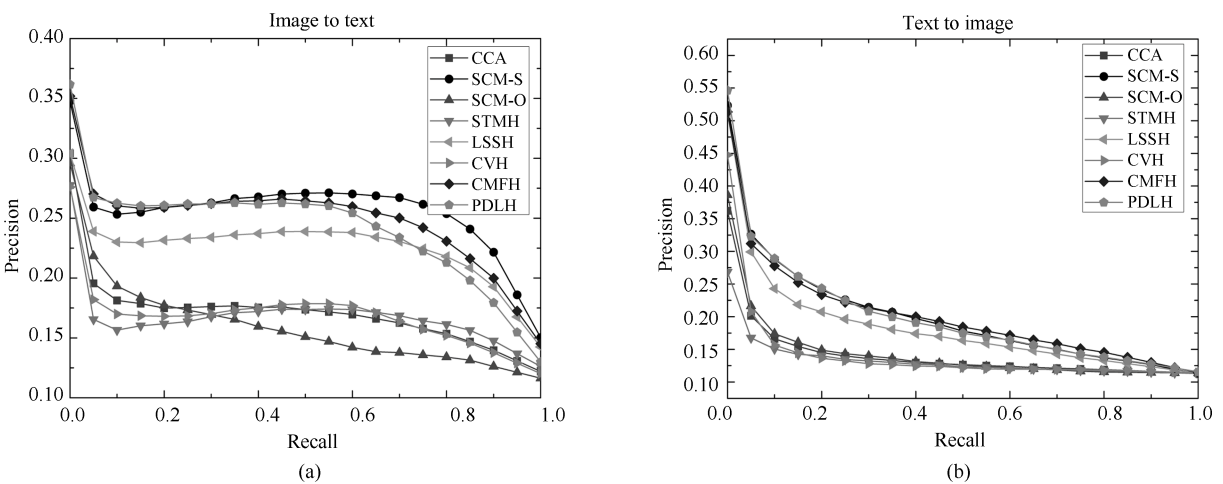


图 3 码长 32 bits 在 Wiki 数据集的 PR 曲线图

Fig. 3 PR curves on Wiki dataset with the code length fixed to 32 bits

表 2 图像检索文本和文本检索图像任务在 NUS-WIDE 数据集上的实验结果 (MAP@200)										
Table 2	MAP results on NUS-WIDE dataset for the tasks of using the image to query texts and vice versa (MAP@200)									
算法	任务	8	16	24	32	任务	8	16	24	32
CCA	图 像 检 索 文 本	0.3445	0.3413	0.3465	0.3424	文	0.3722	0.3620	0.3731	0.3562
CVH		0.3395	0.3435	0.3440	0.3357	本	0.3676	0.3706	0.3620	0.3481
SCM-O		0.3687	0.3580	0.3567	0.3501	检	0.4205	0.4023	0.3866	0.3977
SCM-S		0.4098	0.4443	0.4413	0.4482	索	0.4828	0.5012	0.5067	0.5222
CMFH		0.3374	0.3586	0.3778	0.3803	图	0.3843	0.3984	0.4093	0.4120
LSSH		0.3465	0.3716	0.3770	0.4073	像	0.3686	0.3736	0.3841	0.4184
STMH		0.3723	0.3922	0.4067	0.4156		0.3979	0.4114	0.4235	0.4322
PDLH-		0.4010	0.4423	0.4478	0.4505		0.4362	0.5003	0.5078	0.5128
PDLH		0.4137	0.4456	0.4530	0.4714		0.4530	0.5034	0.5135	0.5172

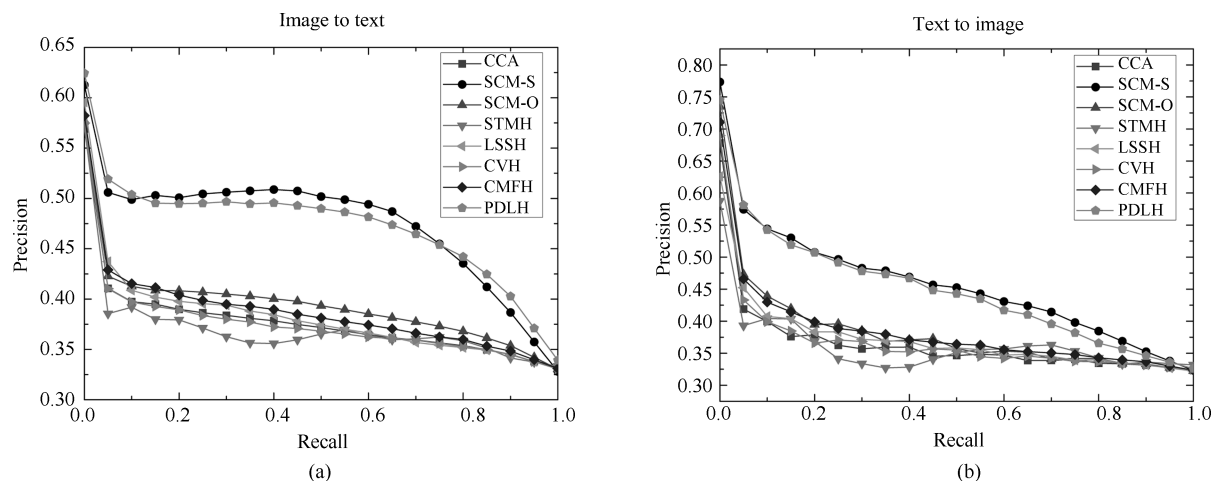


图 4 码长 16 bits 在 NUS-WIDE 数据集的 PR 曲线图

Fig. 4 PR curves on NUS-WIDE dataset with the code length fixed to 16 bits

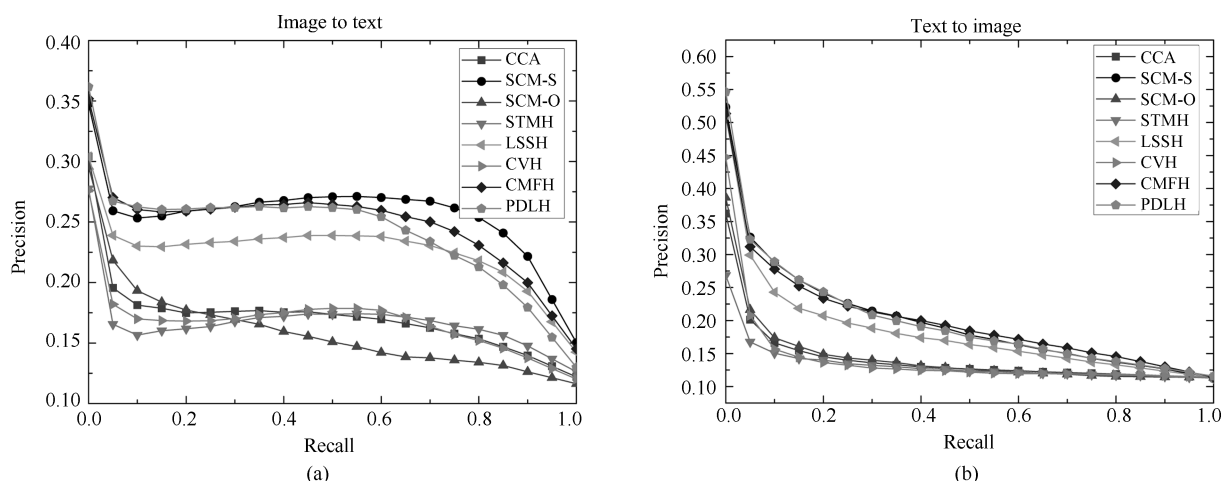


图 5 码长 32 bits 在 NUS-WIDE 数据集的 PR 曲线图

Fig. 5 PR curves on NUS-WIDE dataset with the code length fixed to 32 bits

表 3 不同数量训练样本的训练时间 (s) 和 MAP 结果

Table 3 The time costs (s) and MAP results with different sizes of training dataset

训练集 大小	训练时间 (s)	文本检索图像 MAP	图像检索文本 MAP
10 000	30.25	0.4839	0.4603
20 000	58.75	0.5466	0.4973
50 000	750.77	0.5643	0.5520
100 000	325.90	0.5719	0.5584
150 000	504.59	0.6028	0.5603

References

- Andoni A, Indyk P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science. Berkeley, USA: IEEE, 2006. 459–468
- Kulis B, Kristen G. Kernelized locality-sensitive hashing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(6): 1092–1104
- Li Wu-Jun, Zhou Zhi-Hua. Learning to hash for big data: current status and future trends. *Chinese Science Bulletin*, 2015, **60**(5–6): 485–490
(李武军, 周志华. 大数据哈希学习: 现状与趋势. 科学通报, 2015, **60**(5–6): 485–490)
- Weiss Y, Torralba A, Fergus R. Spectral hashing. In: Proceedings of the 22nd Annual Conference on Neural Information Processing Systems. British Columbia, Canada: MIT, 2008. 1753–1760
- Liu W, Wang J, Ji R R, Jiang Y G, Chang S F. Supervised hashing with kernels. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, Rhode Island, USA: IEEE, 2012. 2074–2081
- Gong Y C, Lazebnik S. Iterative quantization: a procrustean approach to learning binary codes. In: Proceedings of the

- 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA: IEEE, 2011. 817–824
- 7 Shen F M, Shen C H, Liu W, Shen H T. Supervised discrete hashing. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 37–45
- 8 Song J K, Yang Y, Huang Z, Shen H T, Hong R C. Multiple feature hashing for real-time large scale near-duplicate video retrieval. In: Proceedings of the 19th ACM International Conference on Multimedia. New York, USA: ACM, 2011. 423–432
- 9 Zhang D, Wang F, Si L. Composite hashing with multiple information sources. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China: ACM, 2011. 225–234
- 10 Xu H, Wang J D, Li Z, Zeng G, Li S P, Yu N H. Complementary hashing for approximate nearest neighbor search. In: Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011. 1631–1638
- 11 Bronstein M M, Bronstein A M, Michel F, Paragios N. Data fusion through cross-modality metric learning using similarity-sensitive hashing. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 3594–3601
- 12 Kumar S, Udupa R. Learning hash functions for cross-view similarity search. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain: AAAI, 2011. 1360–1366
- 13 Ding G G, Guo Y C, Zhou J L. Collective matrix factorization hashing for multimodal data. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 2083–2080
- 14 Zhou J L, Ding G G, Guo Y C. Latent semantic sparse hashing for cross-modal similarity search. In: Proceedings of the 37th ACM SIGIR Conference on Research and Development in Information Retrieval. Gold Coast, Australia: ACM, 2014. 415–424
- 15 Zhuang Y T, Wang Y F, Wu F, Zhang Y, Lu W M. Supervised coupled dictionary learning with group structures for multi-modal retrieval. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI, 2013. 1070–1076
- 16 Zhen Y, Yeung D Y. A probabilistic model for multimodal hash function learning. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China: ACM, 2012. 940–948
- 17 Rafailidis D, Crestani F. Cluster-based joint matrix factorization hashing for cross-modal retrieval. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy: ACM, 2016. 781–784
- 18 Hotelling H. Relations between two sets of variates. *Biometrika*, 1936, **28**(3–4): 321–377
- 19 Zhang D Q, Li W J. Large-scale supervised multimodal hashing with semantic correlation maximization. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec, Canada: AAAI, 2014. 2177–2183
- 20 Lian Qiu-Sheng, Shi Bao-Shun, Chen Shu-Zhen. Research advances on dictionary learning models, algorithms and applications. *Acta Automatica Sinica*, 2015, **41**(2): 240–260 (练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展. *自动化学报*, 2015, **41**(2): 240–260)
- 21 Chen Si-Bao, Zhao Ling, Luo Bin. Locality preserving based kernel dictionary learning for sparse representation. *Acta Automatica Sinica*, 2014, **40**(10): 2295–2305 (陈思宝, 赵令, 罗斌. 基于局部保持的核稀疏表示字典学习. *自动化学报*, 2014, **40**(10): 2295–2305)
- 22 Yan Y, Yang Y, Shen H Q, Meng D Y, Liu G W, Hauptmann A, Sebe N. Complex event detection via event oriented dictionary learning. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA: AAAI, 2015. 3841–3847
- 23 Huang Dan-Dan, Sun Yi. Tracking via multitask discriminative local joint sparse appearance model. *Acta Automatica Sinica*, 2016, **42**(3): 402–415 (黄丹丹, 孙怡. 基于判别性局部联合稀疏模型的多任务跟踪. *自动化学报*, 2016, **42**(3): 402–415)
- 24 Sun X X, Nasrabadi N M, Tran T D. Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, **53**(8): 4457–4471
- 25 Ma Ming-Lang, He Xiao-Hai, Teng Qi-Zhi, Chen Hong-Gang, Qing Lin-Bo. Fingerprint image compression algorithm via adaptive sparse transformation. *Acta Automatica Sinica*, 2016, **42**(8): 1274–1284 (马名浪, 何小海, 滕奇志, 陈洪刚, 卿琳波. 基于自适应稀疏变换的指纹图像压缩. *自动化学报*, 2016, **42**(8): 1274–1284)
- 26 Zheng Si-Long, Li Yuan-Xiang, Wei Xian, Peng Xi-Shuai. Nonlinear dimensionality reduction based on dictionary learning. *Acta Automatica Sinica*, 2016, **42**(7): 1065–1076 (郑思龙, 李元祥, 魏宪, 彭希帅. 基于字典学习的非线性降维方法. *自动化学报*, 2016, **42**(7): 1065–1076)
- 27 Gu S H, Zhang L, Zuo W M, Feng X C. Projective dictionary pair learning for pattern classification. In: Proceedings of the 2014 Advances in Neural Information Processing Systems. Montréal, Canada: MIT, 2014. 793–801
- 28 Guo J, Guo Y Q, Kong X W, He R. Discriminative analysis dictionary learning. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI, 2016. 1617–1623
- 29 Rasiwasia N, Pereira J C, Coviello E, Doyle G, Lanckriet G R G, Levy R, Vasconcelos N. A new approach to cross-modal multimedia retrieval. In: Proceedings of the 18th ACM International Conference on Multimedia. New York, USA: ACM, 2010. 251–260
- 30 Chua T S, Tang J H, Hong R C, Li H J, Luo Z P, Zheng Y T. NUS-WIDE: a real-world web image database from national university of Singapore. In: Proceedings of the 2009 ACM International Conference on Image and Video Retrieval. Santorini Island, Greece: ACM, 2009. Article No. 48

- 31 Wang D, Gao X B, Wang X M, He L H. Semantic topic multimodal hashing for cross-media retrieval. In: Proceedings of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI, 2015. 3890–3896



姚 涛 大连理工大学信息与通信工程学院博士研究生. 主要研究方向为多媒体检索, 计算机视觉与模式识别.

E-mail: yaotaoedu@mail.dlut.edu.cn

(**YAO Tao** Ph.D. candidate at the School of Information and Communication Engineering, Dalian University of Technology. His research interest covers

multimedia retrieval, computer vision, and machine learning.)

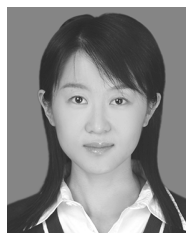


孔祥维 浙江大学数据科学与管理工程学系教授. 2003 年获得大连理工大学管理科学与工程专业博士学位. 2006~2007 年美国普渡大学访问学者. 主要研究方向为人工智能和商务分析, 大数据分析, 跨媒体检索和安全. 本文通信作者.

E-mail: kongxiangwei@zju.edu.cn

(**KONG Xiang-Wei** Professor at the Department of Data Science and Engineering Management, Zhejiang University. She received her Ph.D. degree in management science and engineering from Dalian University of Technology, in 2003. She is a visiting researcher at Purdue University, USA, from 2006 to 2007. Her research interest covers artificial intelligence and busi-

ness analysis, big data analysis, cross-modal retrieval and security. Corresponding author of this paper.)



付海燕 大连理工大学信息与通信工程学院副教授. 2014 年获得大连理工大学博士学位. 主要研究方向为图像检索和计算机视觉. E-mail: fuhay@dlut.edu.cn

(**FU Hai-Yan** Associate professor at the School of Information and Communication Engineering, Dalian University of Technology. She received her Ph.D.

degree from Dalian University of Technology in 2014. Her research interest covers image retrieval and computer vision.)



TIAN Qi 美国德克萨斯大学圣安东尼奥分校计算机科学学院教授. IEEE Fellow. 2002 年获得伊利诺伊大学厄巴纳-香槟分校电子与计算工程博士学位. 主要研究方向为多媒体信息检索, 模式识别和计算机视觉.

E-mail: qitian@cs.utsa.edu

(**TIAN Qi** Professor in the Department of Computer Science at the University of Texas at San Antonio, USA. IEEE Fellow. He received Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana-Champaign in 2002. His research interest covers multimedia information retrieval, machine learning, and computer vision.)