

# 基于标记增强的离散跨模态哈希方法\*

王永欣<sup>1,2</sup>, 田洁茹<sup>2</sup>, 陈振铎<sup>2</sup>, 罗昕<sup>2</sup>, 许信顺<sup>2</sup>

<sup>1</sup>(山东建筑大学 计算机科学与技术学院, 山东 济南 250101)

<sup>2</sup>(山东大学 软件学院, 山东 济南 250101)

通信作者: 许信顺, E-mail: [xuxinshun@sdu.edu.cn](mailto:xuxinshun@sdu.edu.cn)



**摘要:** 跨模态哈希通过将不同模态的数据映射为同一空间中更紧凑的哈希码, 可以大大提升跨模态检索的效率. 然而现有跨模态哈希方法通常使用二元相似性矩阵, 不能准确描述样本间的语义相似关系, 并且存在平方复杂度问题. 为了更好地挖掘数据间的语义相似关系, 提出了一个基于标记增强的离散跨模态哈希方法. 首先借助迁移学习的先验知识生成样本的标记分布, 然后通过标记分布构建描述度更强的语义相似性矩阵, 再通过一个高效的离散优化算法生成哈希码, 避免了量化误差问题. 最后, 在两个基准数据集上的实验结果验证了所提方法在跨模态检索任务上的有效性.

**关键词:** 跨模态检索; 哈希; 标记增强; 迁移学习; 离散优化

**中图法分类号:** TP391

中文引用格式: 王永欣, 田洁茹, 陈振铎, 罗昕, 许信顺. 基于标记增强的离散跨模态哈希方法. 软件学报, 2023, 34(7): 3438–3450. <http://www.jos.org.cn/1000-9825/6536.htm>

英文引用格式: Wang YX, Tian JR, Chen ZD, Luo X, Xu XS. Label Enhancement Based Discrete Cross-modal Hashing Method. Ruan Jian Xue Bao/Journal of Software, 2023, 34(7): 3438–3450 (in Chinese). <http://www.jos.org.cn/1000-9825/6536.htm>

## Label Enhancement Based Discrete Cross-modal Hashing Method

WANG Yong-Xin<sup>1,2</sup>, TIAN Jie-Ru<sup>2</sup>, CHEN Zhen-Duo<sup>2</sup>, LUO Xin<sup>2</sup>, XU Xin-Shun<sup>2</sup>

<sup>1</sup>(School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China)

<sup>2</sup>(School of Software, Shandong University, Jinan 250101, China)

**Abstract:** Cross-modal hashing can greatly improve the efficiency of cross-modal retrieval by mapping data of different modalities into more compact hash codes. Nevertheless, existing cross-modal hashing methods usually use a binary similarity matrix, which cannot accurately describe the semantic similarity relationships between samples and suffer from the squared complexity problem. In order to better mine the semantic similarity relationships of data, this study presents a label enhancement based discrete cross-modal hashing method (LEDCH). It first leverages the prior knowledge of transfer learning to generate the label distribution of samples, then constructs a stronger similarity matrix through the label distribution, and generates the hash codes by an efficient discrete optimization algorithm with a small quantization error. Finally, experimental results on two benchmark datasets validate the effectiveness of the proposed method on cross-modal retrieval tasks.

**Key words:** cross-modal retrieval; hashing; label enhancement; transfer learning; discrete optimization

近年来, 随着互联网上的多媒体数据呈爆炸式增长, 传统的信息检索算法面临巨大的挑战. 为了提升检索效率, 基于哈希学习的数据表示方法成为机器学习和数据检索领域的热点<sup>[1]</sup>. 其基本思想是把高维实值数据映射为海明空间中的紧凑二进制编码; 然后通过计算二进制编码的海明距离来得到样本之间的相似性, 并返回查询结

\* 基金项目: 国家自然科学基金 (62172256, 61872428, 61991411); 山东省重点研发计划 (2019JZZY010127); 山东省自然科学基金 (ZR2019ZD06, ZR2020QF036)

收稿时间: 2021-02-06; 修改时间: 2021-05-27; 采用时间: 2021-11-16; jos 在线出版时间: 2022-09-30

CNKI 网络首发时间: 2022-11-15

果<sup>[2]</sup>. 因为该类方法把数据压缩为短的二进制编码, 所以可以大幅降低数据存储空间; 另外, 在进行二进制编码间的相似性计算时, 计算机可以执行按位异或运算得到海明距离, 所以其计算效率非常高. 基于以上的优越性, 越来越多的研究者从不同角度设计实现了不同的哈希方法<sup>[3-8]</sup>.

已有的方法大多是针对图像检索任务设计的. 然而, 数据通常是由诸如图像、文本、音频、视频等多种异构模态来表示的<sup>[9,10]</sup>. 用户往往希望通过一种模态检索与它相关的其他模态, 例如使用文本检索图像. 因此, 跨模态哈希应运而生, 并成为多媒体检索领域的热点<sup>[11-15]</sup>. 根据是否利用监督信息, 现有的跨模态哈希方法大致可以分为无监督方法和监督方法. 一般来讲, 监督方法能够取得更好的性能. 在现有的监督跨模态哈希方法中, 成对相似性保持是一种常用的策略. 它旨在使哈希码的内积与成对相似性之间的差异最小化, 从而将样本间的相似性嵌入到哈希码中. 但是这种方法往往存在两个问题: 1) 传统的  $(-1, 1)$  二值相似性矩阵丢弃了大量的监督信息, 不能准确描述样本间的相关性, 并且它的计算复杂度通常是  $O(n^2)$ . 另外, 数据的监督信息通常是  $(0, 1)$  标记向量, 这种逻辑标记不能反映每个类别对样本的描述度, 即不能反映数据的真实类别分布; 2) 哈希码的离散优化是一个 NP-hard 问题, 如果采用松弛策略会带来较大的量化误差. 因此, 在这种相似性保持模型中, 如何设计一个更精准的相似度并避免  $O(n^2)$  复杂度, 以及如何设计一个离散优化算法避免量化误差, 是提升检索性能的关键.

为了解决上述问题, 本文提出了一种基于标记增强的离散跨模态哈希方法 (label enhancement discrete cross-modal hashing, LEDCH). 它首先借助迁移学习中的先验知识对数据的逻辑标记进行标记增强, 生成一种更具表达能力的标记分布; 然后利用该标记分布生成更精准的样本间相似性, 这种相似性不仅可以更好地表征样本间的相关性, 而且可以避免平方复杂度问题. 此外, 它离散地生成哈希码, 避免了量化误差问题.

本文的主要贡献包括 3 个方面.

(1) 本文提出了一种基于标记增强的跨模态哈希方法 LEDCH, 它通过基于迁移学习的标记增强生成样本的标记分布, 然后构造一个更精准的相似性矩阵. 它的时间和空间复杂度与训练数据集的规模线性相关. 因此, LEDCH 能够从准确率和效率两方面提升检索性能;

(2) 本文提出了一个高效的离散优化算法来求解哈希码的离散约束问题, 减小了量化误差;

(3) 在两个基准数据集上的实验结果表明了 LEDCH 在跨模态检索任务上的有效性.

## 1 相关工作

为了实现高效跨模态检索, 一些跨模态哈希检索方法被提出. 根据是否应用监督信息, 它们可以被划分为无监督方法和监督方法. 其中, 无监督方法通过保持原始特征的模态内或模态间的相关性来学习哈希码和哈希函数. 例如, 协同矩阵分解哈希 (collective matrix factorization hashing, CMFH)<sup>[16]</sup>首次提出利用协同矩阵分解来学习不同模态的公共子空间; 融合相似性哈希 (fusion similarity hashing, FSH)<sup>[17]</sup>将不同模态的融合相似性嵌入到哈希码中; 稳健灵活的离散哈希 (robust and flexible discrete hashing, RFDH)<sup>[18]</sup>通过离散协同矩阵分解来生成哈希码, 并且通过一个分类模型来学习哈希函数; 协同重构嵌入 (collective reconstructive embeddings, CRE)<sup>[19]</sup>对不同的模态应用不同的距离度量来解决多模态异构鸿沟问题. 这些无监督方法取得了一定的效果; 但是, 它们不能利用监督信息, 容易产生语义鸿沟问题.

监督方法进一步利用监督信息来生成更具判别性的哈希码, 通常可以取得更好的检索性能. 例如, 监督核哈希 (supervised hashing with kernels, KSH)<sup>[5]</sup>使相似样本的哈希码间的海明距离最小, 而不相似样本的哈希码间的海明距离最大, 从而将数据的成对相似性嵌入到哈希码中; 语义嵌入哈希 (semantics-preserving hashing, SePH)<sup>[20]</sup>将数据的成对相似性矩阵转换为一个概率分布, 然后通过最小化信息散度将其嵌入到哈希码中; 半松弛监督哈希 (semi-relaxation supervised hashing, SRS<sub>H</sub>)<sup>[21]</sup>通过分解一个成对相似性矩阵来学习哈希码和哈希函数, 并应用半松弛策略求解离散约束问题; 广义语义嵌入哈希 (generalized semantic preserving hashing, GSPH)<sup>[22]</sup>采用一个两步哈希策略, 首先最小化成对相似性矩阵与两个松弛变量内积的差异来学习哈希码, 然后学习数据到哈希码的映射作为哈希函数. 但是这些方法由于应用尺寸为  $n \times n$  的二值成对相似性矩阵而存在平方复杂度问题. 为了解决这个问题, 语义相关性最大化 (semantic correlation maximization, SCM)<sup>[23]</sup>直接将语义标签集成到大规模数据的哈希学习过程

中; 判别的跨模态哈希 (discriminative cross-modal hashing, DCH)<sup>[24]</sup>提出将哈希学习看成一个分类问题, 构建哈希码到标记矩阵的映射; 可扩展的离散矩阵分解哈希 (scalable discrete matrix factorization hashing, SCRATCH)<sup>[25]</sup>应用矩阵分解学习标签矩阵和不同模态的公共子空间, 然后通过该子空间学习哈希码; 离散潜变量哈希 (discrete latent factor hashing, DLFH)<sup>[26]</sup>应用离散潜变量模型将监督信息离散地嵌入到哈希码中. 但是, 上述所有监督跨模态哈希方法都没有考虑样本的真实语义分布. 无论是逻辑标记还是二值成对相似性, 都是一种粗粒度的语义信息. 我们认为更精准和监督信息对于生成更精细的检索结果具有重要意义.

从另一个角度, 根据是否对哈希码的离散约束进行松弛, 哈希方法又可以划分为离散方法和松弛方法<sup>[27]</sup>. 由于哈希码的特殊性, 导致直接求解离散约束是一个 NP-hard 问题. 通常的做法是对其进行松弛, 首先求得一个哈希码的实值近似, 然后再二值化为哈希码<sup>[22,28]</sup>. 显然, 这种方式会导致较大的量化误差, 从而降低哈希码的质量. 为了解决这个问题, 文献 [3] 通过应用一个旋转矩阵来减小实值近似与哈希码之间的量化误差. 文献 [4] 应用符号化梯度上升 (signed gradient ascent, SGA) 生成离散哈希码. 文献 [24,29] 应用离散循环坐标 (discrete cyclic coordinate, DCC) 来按位更新哈希码. 但是, 上述优化方法仍然存在误差大、计算效率低等问题.

近年来, 也有一些深度哈希方法应用深度神经网络来学习数据的哈希函数<sup>[30-33]</sup>. 它们的优势是可以直接处理原始数据, 但缺陷是它们的计算复杂度通常都很高, 往往需要几小时甚至几十小时来完成训练; 另外, 它们通常专注于设计深度网络的结构, 而忽略了目标函数和优化算法的重要性. 与之相反, 浅层哈希方法具有训练速度快、可解释性强、损失函数多变、优化灵活等特点, 在实际应用上更具有优势<sup>[34]</sup>. 因此, 本文主要研究浅层哈希方法, 重点在于设计良好的损失函数来生成高质量的哈希码和设计高效的离散优化算法来减小量化误差.

本文方法 LEDCH 首次利用标记增强来生成更精准和监督信息, 并且借助迁移学习引入先验知识, 实现了监督信息的增强; 另外, LEDCH 通过高效的离散优化算法, 减小了量化误差, 提高了训练效率, 可以扩展到大数据集. 在基准数据集上的实验结果表明 LEDCH 在准确率和效率方面均能取得比现有流行方法更好的结果.

## 2 基于标记增强的跨模态哈希

本文方法的框架图如图 1 所示, 分为标记增强、哈希码学习和哈希函数学习 3 部分. 其中, 标记增强属于数据预处理过程, 它通过迁移学习生成标记分布; 哈希码学习和哈希函数学习是两步哈希的两个阶段, 哈希码学习通过内积适应将标记分布相似性嵌入到哈希码中, 哈希函数学习应用上一步生成的哈希码作为监督信息, 通过线性回归学习特征到哈希码的映射函数.

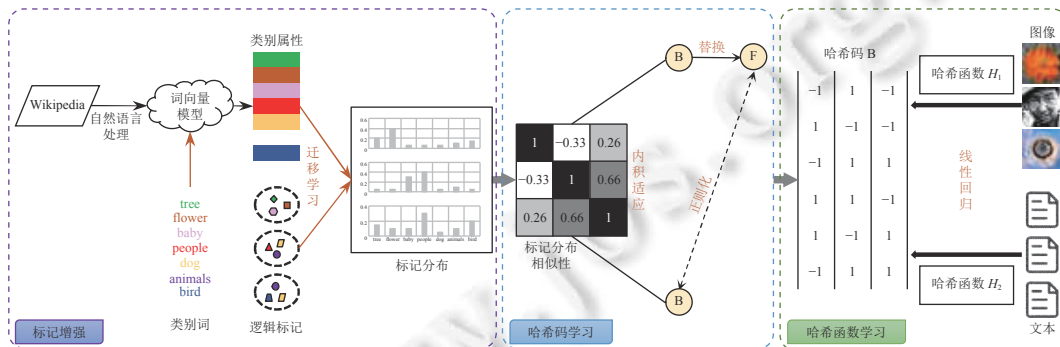


图 1 标记增强离散跨模态哈希的框架

### 2.1 符号定义

$X_{ik}$ 、 $X_j$ 、 $X_{ij}$  分别表示矩阵  $X$  的第  $i$  行、第  $j$  列和第  $(i, j)$  元素.  $I$ 、 $1$ 、 $0$  分别表示单位矩阵、全一矩阵和全零矩阵.  $\|\cdot\|$  表示矩阵的  $\ell_F$  范数或者向量的  $\ell_2$  范数.  $tr(\cdot)$  表示方阵的迹.  $sgn(\cdot)$  是一个符号化函数, 当输入为非负值时返回 1, 否则返回 -1.

假设训练数据包含  $p$  个模态,  $X^{(l)} \in \mathbb{R}^{d_l \times n}$  是第  $l \in \{1, 2, \dots, p\}$  模态的训练样本, 其中  $n$  是样本数,  $d_l$  是特征的维度.  $L \in \{0, 1\}^{c \times n}$  表示训练数据的逻辑标签矩阵, 其中  $c$  是类别数. 当第  $j$  样本属于第  $i$  类时,  $L_{ij} = 1$ , 否则  $L_{ij} = 0$ .  $D \in \mathbb{R}^{c \times n}$  表示对逻辑标记  $L$  进行标记增强生成的标记分布.  $B \in \{-1, 1\}^{r \times n}$  表示要学习的训练集的哈希码,  $r$  是哈希码的位数.  $H_l(\cdot)$  表示将第  $l$  模态的查询样本映射为哈希码的哈希函数. 表 1 汇总了本文使用的主要符号.

表 1 LEDCH 中的主要符号

符号	解释
$X^{(l)} \in \mathbb{R}^{d_l \times n}, l \in \{1, 2, \dots, p\}$	第 $l$ 模态的训练样本
$L \in \{0, 1\}^{c \times n}$	训练数据的逻辑标签矩阵
$D \in \mathbb{R}^{c \times n}$	对逻辑标记进行标记增强生成的标记分布
$B \in \{-1, 1\}^{r \times n}$	待学习的训练集的公共哈希码
$H_l(\cdot)$	将第 $l$ 模态的查询样本映射为哈希码的哈希函数
$A \in \mathbb{R}^{c \times k}$	类别属性
$S \in [-1, 1]^{n \times n}$	标记分布相似性矩阵
$W^{(l)} \in \mathbb{R}^{r \times d_l}$	第 $l$ 模态查询样本的投影矩阵
$p$	训练数据包含的模态数
$n$	训练样本的个数
$d_l$	第 $l$ 模态训练样本的特征维度
$c$	训练样本的类别数
$r$	待学习哈希码的位数
$k$	类别属性的特征维度

## 2.2 标记增强

### 2.2.1 迁移学习

标记增强是一种从数据的逻辑标记中复原标记分布的策略<sup>[35,36]</sup>. 它通常采用的做法是利用特征空间的拓扑结构对标记进行增强, 在人脸识别、年龄预测等任务中有不错的表现. 但是, 对自然场景数据来说, 特征空间与标记空间有较大的语义鸿沟, 采用拓扑结构的方式可能反而降低模型性能.

从目标上看, 我们认为标记增强与零样本学习有一些相似之处. 标记增强是通过已知的逻辑标记、训练数据和一些其他知识学习未知的标记分布; 零样本学习是通过已知的逻辑标记、训练数据和一些其他知识学习并预测未知类的样本. 它们两者都有信息的增强过程. 在零样本学习中, 往往通过引入类别属性来链接高层语义与底层特征. 其中一种广泛采用的策略就是利用迁移学习, 从一个巨大的语料库 Wikipedia 数据集中学习词向量作为类别属性<sup>[37]</sup>. 通过这种方式学习到的类别属性, 具有较强的表达能力, 并且避免了人为设计.

受此启发, 我们认为标记增强也可以通过应用迁移学习中的类别属性作为先验知识来促进学习, 因为先验知识通常隐含一些标记分布的天然性质. 具体来说, 我们首先根据样本语义得到对应的类别词  $o = \{o_1; o_2; \dots; o_c\}$ , 然后将 Wikipedia 上完成自然语言处理预训练的词向量模型<sup>[38]</sup>迁移到本文任务中, 将类别词输入词向量模型, 将输出作为这几个类的类别属性  $A \in \mathbb{R}^{c \times k}$ . 其中,  $k$  是类别属性的维度. 一个类别往往包含一些特性, 比如颜色、形状和共存关系等. 词向量基于自然语言处理技术将词表示为向量, 理论上可以将类别词包含的特性编码为类别属性向量.

在此基础上, 将该类别属性迁移到本文的标记增强中. 要生成的标记分布应该满足如下两个条件: 1) 标记分布  $D$  中应包含类别属性  $A$  的信息; 2) 标记分布  $D$  应与逻辑标记  $L$  尽可能相似. 为了实现上述目标, 我们定义如下目标函数:

$$\min_{R, P, D} \|L - RD\|^2 + \alpha \|P^T L - A^T D\|^2 + \theta \|P\|^2, \text{ s.t. } R^T R = I \quad (1)$$

其中,  $\alpha$  是平衡参数,  $\theta$  是正则化项的惩罚系数;  $R \in \mathfrak{R}^{c \times c}$  是正交旋转矩阵,  $P \in \mathfrak{R}^{c \times k}$  是投影矩阵. 公式 (1) 的第 1 项通过引入一个旋转矩阵  $R$  使标记分布  $D$  与逻辑标记  $L$  共享尽可能多的信息; 第 2 项假设存在一个  $k$  维的类别表示空间, 它通过对逻辑标记  $L$  进行投影得到, 即  $P^T L$ . 同时, 该类别表示空间可以分解为语义类别空间, 即标记分布  $D$  与语义类别属性  $A$  的乘积.

### 2.2.2 问题求解

为了解公式 (1), 我们提出一个 3 步迭代策略:

(1) 更新  $R$ . 固定哈希码  $B$  和投影矩阵  $P$  不变, 公式 (1) 可以转换为如下关于  $R$  的子问题:

$$\min_R \|L - RD\|^2, \text{ s.t. } R^T R = I \quad (2)$$

公式 (2) 是一个典型的正交普鲁克问题<sup>[3]</sup>. 对  $LD^T$  进行奇异值分解可以得到  $T\Omega\hat{T}^T$ . 于是, 公式 (2) 的最优解是:

$$R = T\hat{T}^T \quad (3)$$

(2) 更新  $P$ . 令公式 (1) 对  $P$  的导数为零, 可以得到  $P$  的最优解为:

$$P = (\alpha LL^T + \theta I)^{-1} \alpha LD^T A \quad (4)$$

(3) 更新  $D$ . 公式 (1) 对  $D$  求导, 并令导数为零, 可得最优的  $D$  为:

$$D = (\alpha AA^T + I)^{-1} (R^T L + \alpha P^T L) \quad (5)$$

迭代上述步骤直到收敛, 然后对  $D$  做如下软归一化:

$$D_{ij} = e^{D_{ij}} / \sum_i e^{D_{ij}} \quad (6)$$

算法 1 给出了标记增强的完整算法流程. 值得注意的是, 标记增强属于数据预处理过程, 可以在哈希学习之前预先计算, 因此不会影响哈希学习的效率. 实际上, 公式 (1) 可以非常高效地通过上述迭代优化策略求解, 复杂度与训练集的规模线性相关.

---

#### 算法 1. 基于迁移学习的标记增强算法.

---

输入: 逻辑标记矩阵  $L$ , 类别属性矩阵  $A$ , 最大迭代次数  $t_1$ , 参数  $\alpha$ 、 $\theta$ .

1. 随机初始化  $P$  和  $D$ ;
2. repeat
3. 用公式 (3) 更新  $R$ ;
4. 用公式 (4) 更新  $P$ ;
5. 用公式 (5) 更新  $D$ ;
6. until convergent
7. 用公式 (6) 软归一化标记分布  $D$ ;

输出: 标记分布  $D$ .

---

### 2.2.3 收敛性证明

算法 1 是收敛的. 为了证明其收敛性, 我们让  $\mathcal{L}_1(R, P, D)$  表示公式 (1) 的目标函数. 由于每一个变量都在自己的子问题中有闭式解, 可得  $\mathcal{L}_1(R^{i+1}, P^{i+1}, D^{i+1}) \leq \mathcal{L}_1(R^{i+1}, P^i, D^i) \leq \mathcal{L}_1(R^i, P^i, D^i) \leq \mathcal{L}_1(R^i, P^i, D^i)$ , 其中  $t_1^*$  表示迭代轮次. 也就是说, 目标函数的值随着迭代是单调递减的, 又由于  $\mathcal{L}_1(R, P, D)$  是 3 个正范数的和, 因此它有下界. 根据文献 [39] 中的有界单调收敛定理, 该优化算法最终会收敛到一个局部最优解.

## 2.3 哈希码学习

我们采用一个两步哈希策略学习哈希码和哈希函数: 首先利用训练数据的监督信息生成哈希码; 然后用学到的哈希码作为监督信息, 训练一个分类器, 生成新样本的哈希函数.

### 2.3.1 相似性保持

为了学习有效的哈希码, 我们应将样本的相似性尽可能地嵌入到海明空间中. 因此, 我们采用广泛使用的内积

适应目标函数:

$$\min_B \|B^T B - rS\|^2, \text{ s.t. } B \in \{-1, 1\}^{r \times n} \quad (7)$$

一般来讲, 大部分方法采用的  $S \in \{-1, 1\}^{n \times n}$  是训练集的成对相似性矩阵<sup>[5]</sup>. 如果第  $(i, j)$  样本对至少共享一个类别标记, 则  $S_{ij} = 1$ , 否则  $S_{ij} = -1$ . 显然, 这种定义二值相似性的方式损失了大量的语义信息, 尤其是对多标记数据. 例如, 对于 3 个相似的样本  $x_1, x_2, x_3$ , 假设它们的逻辑标记分别为  $l_1 = (1; 1; 1)$ 、 $l_2 = (1; 1; 0)$ 、 $l_3 = (1; 1; 1)$ . 显然  $l_1$  与  $l_2$  的差异要大于  $l_1$  与  $l_3$  的差异. 然而在上述的相似性矩阵中  $S_{12} = S_{13} = 1$ . 另外, 尽管看起来  $l_2$  与  $l_1$  的差异等于  $l_2$  与  $l_3$  的差异, 它们的标记分布可能是  $d_1 = (0.8; 0.1; 0.1)$ 、 $d_2 = (0.9; 0.1; 0)$ 、 $d_3 = (0.1; 0.1; 0.8)$ , 即样本中包含某个类的概率是不同的. 显然  $d_2$  与  $d_1$  的差异小于  $d_2$  与  $d_3$  的差异. 因此即使拥有相同逻辑标记的样本, 也会有不同的语义相似性.

考虑到这些问题, 我们应用第 2.2 节生成的标记分布来构造更细粒度的相似性矩阵, 称之为标记分布相似性. 它的定义如下:

$$S = 2\bar{D}^T \bar{D} - 11^T \quad (8)$$

其中,  $\bar{D}_i = D_i / \|D_i\|$  是按列  $\ell_2$  范数归一化的标记分布矩阵. 值得注意的是, 在实际运算中并不需要真正生成  $S$ . 直接应用公式 (8) 等号右侧参与运算, 可以避免平方复杂度问题, 并且对大规模数据是可扩展的.

### 2.3.2 高效离散优化

由于哈希码的离散约束, 直接求解公式 (7) 是一个 NP-hard 问题. 为了实现高效优化, 我们引入一个中间变量  $F \in \mathcal{R}^{r \times n}$  代替对称内积运算中的一个哈希码  $B$ . 另外, 为了避免  $F$  偏移, 我们对  $F$  应用正交和平衡约束, 并用一个正则化项使  $F$  与  $B$  之间的差异最小. 最终, 公式 (7) 转变为如下形式:

$$\min_{F, B} \|B^T F - rS\|^2 + \omega \|B - F\|^2, \text{ s.t. } B \in \{-1, 1\}^{r \times n}, FF^T = nI, F1 = 0 \quad (9)$$

其中,  $\omega$  是正则化项的惩罚系数.

公式 (9) 可以通过选择性迭代更新实现高效求解, 具体的步骤描述如下.

(1) 更新  $F$ . 固定  $B$ , 公式 (9) 转化为如下关于  $F$  的子问题:

$$\min_F \|B^T F - rS\|^2 + \omega \|B - F\|^2, \text{ s.t. } FF^T = nI, F1 = 0 \quad (10)$$

由于约束  $FF^T = nI$  和  $B \in \{-1, 1\}^{r \times n}$ , 公式 (10) 可以等价如下矩阵迹的形式:

$$\max_F \text{tr}((rBS + \omega B)F^T), \text{ s.t. } FF^T = nI, F1 = 0 \quad (11)$$

为了求解公式 (11), 我们定义  $J = I - \frac{1}{n}11^T$ , 并令  $Z = rBS + \omega B$ . 根据公式 (8) 对相似性矩阵的定义, 有:

$$Z = 2r(B\bar{D}^T)\bar{D} - r(B1)1^T + \omega B \quad (12)$$

然后对  $ZJZ^T$  做矩阵分解可得:

$$ZJZ^T = [V \quad \hat{V}] \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} [V \quad \hat{V}]^T \quad (13)$$

其中,  $\Sigma$  是由  $r' \leq r$  个正特征值组成的对角矩阵,  $V$  是对应的特征向量组成的矩阵,  $\hat{V}$  是其余的零特征值对应的特征向量组成的矩阵. 对  $\hat{V}$  做施密特正交化得到正交矩阵  $\tilde{V} \in \mathcal{R}^{r \times (r-r')}$ , 对一个随机矩阵做施密特正交化得到一个随机正交矩阵  $\tilde{U} \in \mathcal{R}^{n \times (r-r')}$ . 定义  $U = JZ^T V \Sigma^{-1/2}$ , 根据文献 [4], 公式 (11) 的解为:

$$F = \sqrt{n} [V \quad \hat{V}] [U \quad \tilde{U}]^T \quad (14)$$

(2) 更新  $B$ . 当  $F$  固定时, 公式 (9) 变为如下形式:

$$\min_B \text{tr}((rFS + \omega F)B^T), \text{ s.t. } B \in \{-1, 1\}^{r \times n} \quad (15)$$

我们容易得到公式 (15) 的最优解为:

$$B = \text{sgn}(rFS + \omega F) \quad (16)$$

根据公式 (8), 有:

$$B = \text{sgn}(2r(F\bar{D}^T)\bar{D} - r(F1)1^T + \omega F) \quad (17)$$

重复上述两个步骤直到收敛, 最终得到的即是训练集的哈希码. 正如前文所述, 相似性矩阵  $S$  并不需要真正生成. 通过公式 (12) 和公式 (17), 相似关系既嵌入到哈希码中, 又避免了平方复杂度问题. 整个哈希码优化过程的时空复杂度是与训练集的规模线性相关的, 从而对大数据是可扩展的.

### 2.3.3 收敛性证明

上述高效离散优化算法是收敛的. 为了证明其收敛性, 首先让  $\mathcal{L}_2(F, B)$  表示公式 (9) 的目标函数. 与第 2.2.3 节的证明类似, 有  $\mathcal{L}_2(F^{t_2+1}, B^{t_2+1}) \leq \mathcal{L}_2(F^{t_2+1}, B^{t_2}) \leq \mathcal{L}_2(F^{t_2}, B^{t_2})$ . 其中  $t_2$  表示迭代轮次. 由于目标函数是两个正范数的和, 而且它的值是随着迭代单调递减的, 根据文献 [39] 的单调有界收敛理论, 可证该优化算法也是收敛的.

## 2.4 哈希函数学习

如前文所述, 本文提出的方法是一个两步哈希方法. 在得到训练集的哈希码之后, 我们需要学习不同模态的新样本的哈希函数. 有许多模型可以实现这一目标<sup>[40-43]</sup>, 例如线性回归、决策树、支持向量机、深度神经网络等. 一般来讲, 越复杂的模型对应越高的准确率以及越高的计算复杂度. 为了平衡准确率和效率, 我们选择线性回归模型. 实际上, 在该步骤应用的模型越简单, 越能证明第 1 步生成哈希码的质量. 具体来说, 我们使用第 1 步生成的哈希码作为监督信息, 训练一个第  $l$  模态的特征到哈希码的线性映射, 目标函数如下:

$$\min_{W^{(l)}} \|B - W^{(l)} X^{(l)}\|^2 + \lambda \|W^{(l)}\|^2 \quad (18)$$

其中,  $\lambda$  是避免过拟合的正则化项的惩罚系数,  $W^{(l)}$  是第  $l$  模态的映射矩阵. 通过对  $W^{(l)}$  求导并令导数为零, 可以得到公式 (18) 的最优解为:

$$W^{(l)} = BX^{(l)\top} (X^{(l)} X^{(l)\top} + \lambda I)^{-1} \quad (19)$$

给定一个第  $l$  模态的查询样本  $q^{(l)}$ , 它的哈希码可以通过如下哈希函数得到:

$$H_l(q^{(l)}) = \text{sgn}(W^{(l)} q^{(l)}) \quad (20)$$

算法 2 给出了哈希学习的完整算法流程, 包括哈希码学习和哈希函数学习两个阶段.

---

### 算法 2. 基于标记增强的离散哈希算法.

---

输入: 训练数据集  $X^{(l)}$  ( $l \in \{1, \dots, p\}$ ), 标记分布  $D$ ; 哈希码长度  $r$ , 最大迭代次数  $t_2$ , 参数  $\omega$ 、 $\lambda$ ;

1. 用符号化的正态分布随机初始化  $B$ ;
2. repeat
3. 用公式 (14) 更新  $F$ ;
4. 用公式 (17) 更新  $B$ ;
5. until convergent
6. 用公式 (19) 计算第  $l$  模态的投影矩阵  $W^{(l)}$ ;
7. 用公式 (20) 计算新样本的哈希码;

输出: 训练样本的哈希码  $B$ , 查询样本的哈希函数  $H_l(q^{(l)})$ .

---

## 3 实验结果

在本节中, 我们主要介绍实验用到的基准数据集、评价标准、对比方法与实现细节、实验结果与分析.

### 3.1 数据集

为了验证 LEDCH 的性能, 我们在 MIRFlickr-25K<sup>[44]</sup> 和 NUS-WIDE<sup>[45]</sup> 两个常用的基准数据集上进行了实验. 具体的数据集设置描述如下.

MIRFlickr-25K 是一个从 Flickr 中采集, 包含 25 000 个图像-文本对的数据集. 它是一个多标记数据集, 每个样本至少属于 24 个类别中的一个. 每幅图片用 512 维的 GIST 特征向量表示, 每个文本用 1 386 维的词袋向量表示. 参考文献 [26], 我们去掉了那些没有标记的样本. 最终剩余 20 015 个样本. 我们随机选取了 2 000 个样本作为查询

集, 其余的作为检索集和训练集.

NUS-WIDE 是一个包含 269 468 个图像和文本的真实世界数据集. 它包含 81 个语义类别, 每个样本至少属于其中一个. 每幅图片用 500 维的 SIFT 特征向量表示, 每个文本用 1 000 维的标签向量表示. 与文献 [26] 类似, 我们也仅选用那些属于前 10 类别的样本, 最终剩余 186 577 个样本对. 我们随机选取 2 000 个作为查询集, 其他的作为检索集和训练集.

### 3.2 评价标准

为了从准确率和效率两方面评价本文提出的 LEDCH 的性能, 我们在图像检索文本和文本检索图像这两个跨模态检索任务上进行测试. 选用的评价指标为检索任务中广泛采用的平均精度均值 (MAP) 和训练时间 (training time). 其中, 平均精度均值是通过计算平均精度然后求均值得到的. 它的值越大表示准确率越高. 如果两个样本至少共享一个标记就认为它们是真实的近邻关系. 训练时间是通过统计算法从开始到结束的运行时间得到的. 它的值越小说明效率越高.

### 3.3 对比方法和实现细节

我们将 LEDCH 与 7 个监督跨模态哈希方法进行了对比实验, 它们分别是语义相关性最大化哈希 (semantic correlation maximization, SCM)<sup>[23]</sup>, 离散跨模态哈希 (discriminative cross-modal hashing, DCH)<sup>[24]</sup>, 快速离散跨模态哈希 (fast discrete cross-modal hashing, FDCH)<sup>[46]</sup>, 可扩展离散矩阵分解哈希 (scalable discrete matrix factorization hashing, SCRATCH)<sup>[25]</sup>, 标记一致矩阵分解哈希 (label consistent matrix factorization hashing, LCMFH)<sup>[28]</sup>, 离散潜在因子哈希 (discrete latent factor hashing, DLFH)<sup>[26]</sup>, 子空间关系学习跨模态哈希 (subspace relation learning for cross-modal hashing, SRLCH)<sup>[47]</sup>. 其中, FDCH 和 SRLCH 是我们根据论文的介绍实现的, 其他对比方法都是根据作者提供的代码实现的, 所有参数均根据论文的建议进行设置. 对于本文的 LEDCH, 标记增强中类别属性的维度  $k$  被设置为 500, 参数  $\alpha$ 、 $\theta$  均被设置为 1; 哈希码学习过程中的参数  $\omega$ 、 $\lambda$  均被设置为 1. 另外, 综合考虑收敛率和优化效率, 两个迭代次数  $t_1$  和  $t_2$  分别被设置为 10 和 5.

### 3.4 实验结果与分析

#### 3.4.1 准确率分析

表 2 展示了本文方法 LEDCH 与所有对比方法在 MIRFlickr-25K 和 NUS-WIDE 两个基准数据集上的 MAP 结果. 最好的 MAP 结果用粗体表示. 我们从表 2 可以得出如下结论.

表 2 LEDCH 和所有对比方法在 MIRFlickr-25K、NUS-WIDE 数据集上的 MAP 分数

任务	方法	MIRFlickr-25K					NUS-WIDE				
		8位	16位	32位	64位	128位	8位	16位	32位	64位	128位
图像检索文本	SCM-seq	0.6456	0.6600	0.6701	0.6570	0.6595	0.4935	0.4546	0.5459	0.5316	0.5305
	DCH	0.7041	0.7108	0.7202	0.7349	0.7613	0.6310	0.6078	0.5715	0.5836	0.5936
	FDCH	0.6561	0.7021	0.7122	0.7138	0.7139	0.5328	0.6136	0.6343	0.6454	0.6421
	SCRATCH	0.7073	0.7094	0.7268	0.7322	0.7377	0.6129	0.6252	0.6404	0.6427	0.6497
	LCMFH	0.6864	0.6966	0.7062	0.7045	0.7148	0.5698	0.5806	0.6221	0.6327	0.6395
	DLFH	0.6971	0.7224	0.7298	0.7418	0.7454	0.5828	0.6253	0.6475	0.6651	0.6712
	SRLCH	0.6652	0.6570	0.6744	0.6854	0.6980	0.5952	0.6390	0.6262	0.6653	0.6654
	LEDCH	<b>0.7467</b>	<b>0.7547</b>	<b>0.7562</b>	<b>0.7648</b>	<b>0.7655</b>	<b>0.6509</b>	<b>0.6579</b>	<b>0.6721</b>	<b>0.6737</b>	<b>0.6761</b>
文本检索图像	SCM-seq	0.6296	0.6414	0.6604	0.6353	0.6390	0.4794	0.4540	0.5262	0.5246	0.5280
	DCH	0.7615	0.7636	0.7650	0.7956	0.8246	0.7644	0.7433	0.6895	0.7065	0.7344
	FDCH	0.7210	0.7683	0.7985	0.7428	0.8090	0.6919	0.7658	0.7875	0.7884	0.8067
	SCRATCH	0.7553	0.7716	0.7830	0.7931	0.8054	0.7348	0.7449	0.7732	0.7823	0.7918
	LCMFH	0.7370	0.7551	0.7755	0.7723	0.7916	0.6631	0.6718	0.7289	0.7453	0.7548
	DLFH	0.7550	0.8076	0.8252	0.8392	0.8464	0.6693	0.7769	0.8042	0.8162	0.8199
	SRLCH	0.6927	0.6912	0.7273	0.7304	0.7535	0.7221	0.7632	0.7635	0.7981	0.8086
	LEDCH	<b>0.8247</b>	<b>0.8395</b>	<b>0.8465</b>	<b>0.8567</b>	<b>0.8582</b>	<b>0.7718</b>	<b>0.7998</b>	<b>0.8163</b>	<b>0.8191</b>	<b>0.8230</b>



(1) LEDCH 在两个数据集的两种跨模态任务上都取得了最好的效果, 表明了它的有效性.

(2) 与 LCMFH、DLFH 等几个最新的方法相比, LEDCH 仍取得了可观的性能提升, 这说明本文所提出的基于迁移学习的标记增强能够挖掘到数据的真实语义分布, 并通过保持标记分布相似性学习到了高质量的哈希码.

(3) 随着哈希码位数的升高, 大多数方法的性能都有一定程度的提升, 揭示了长哈希码能够包含更多的相似信息. 但是 LEDCH 在短码上仍能取得更好的性能, 表明它具有较强的信息嵌入能力, 从而能够学习到更具有判别力的哈希码.

(4) 大多数方法在文本检索图像任务上比在图像检索文本任务上表现得更好. 一个可能的原因是, 文本特征比人为设计的图像特性更能描述样本的内容.

### 3.4.2 效率分析

表 3 展示了所有对比方法和 LEDCH 的训练时间. 最好的结果用粗体表示. 从表 3 我们可以看出:

(1) LEDCH 在两个数据集上的训练时间都是最短的, 证明了它的高效性.

(2) 尽管所有方法都声称线性复杂度, 它们的训练时间仍然千差万别. 因为不同算法对哈希码长度、特征维度、迭代次数等的敏感度不一样.

表 3 LEDCH 和所有对比方法在 MIRFlickr-25K、NUS-WIDE 数据集上的训练时间 (s)

方法	MIRFlickr-25K					NUS-WIDE				
	8位	16位	32位	64位	128位	8位	16位	32位	64位	128位
SCM-seq	13.46	22.57	44.28	57.24	99.33	10.21	11.64	22.03	44.09	97.85
DCH	6.64	6.18	11.96	17.76	60.23	37.80	40.08	78.21	193.36	814.99
FDCH	13.65	13.42	13.52	13.78	14.90	82.42	86.91	88.12	89.42	86.83
SCRATCH	2.25	2.96	5.54	7.17	8.95	26.39	30.88	37.86	49.71	81.75
LCMFH	8.99	9.75	10.84	10.56	14.46	45.57	44.32	50.92	63.00	75.69
DLFH	4.74	4.84	21.69	37.60	241.48	13.39	33.55	160.72	611.47	2498.0
SRLCH	14.65	14.43	15.09	15.32	15.93	130.43	131.71	133.09	122.60	140.22
LEDCH	<b>1.67</b>	<b>1.70</b>	<b>1.72</b>	<b>1.61</b>	<b>2.38</b>	<b>8.71</b>	<b>8.84</b>	<b>9.57</b>	<b>13.79</b>	<b>15.87</b>

### 3.4.3 消融分析

为了进一步证明本文提出的标记增强的有效性, 我们还进行了消融实验. 表 4 给出了 LEDCH 与两个变种的 MAP 结果. 其中, LEDCH-S 表示采用传统二值相似性矩阵的方法. 由于这种二值相似性矩阵存在平方复杂度问题, 无法使用全部的训练样本参与训练, 我们随机选取 10 000 个样本用于 LEDCH-S 的训练. 尽管 LEDCH 可以高效地使用全部训练数据, 为了更好地对比, 我们用 LEDCH-10K 表示 LEDCH 只使用 10 000 个训练样本. 从表格中我们有如下发现:

(1) 比较 LEDCH-S 与 LEDCH-10K 的结果, 我们可以看出本文提出的标记分布相似性能够明显提升检索准确率.

(2) 对比 LEDCH-10K 与 LEDCH, 我们可以看出应用全部训练集能够明显提升准确率, 这也说明了解决平方复杂度问题的重要意义.

表 4 LEDCH 和它的变形在 MIRFlickr-25K、NUS-WIDE 数据集上的 MAP 分数

任务	方法	MIRFlickr-25K					NUS-WIDE				
		8位	16位	32位	64位	128位	8位	16位	32位	64位	128位
图像检索文本	LEDCH-S	0.6640	0.6821	0.7084	0.7144	0.7185	0.5095	0.5253	0.5321	0.5409	0.5450
	LEDCH-10K	0.6938	0.7039	0.7137	0.7224	0.7279	0.5206	0.5370	0.5410	0.5478	0.5535
	LEDCH	<b>0.7467</b>	<b>0.7547</b>	<b>0.7562</b>	<b>0.7648</b>	<b>0.7655</b>	<b>0.6509</b>	<b>0.6579</b>	<b>0.6721</b>	<b>0.6737</b>	<b>0.6761</b>
文本检索图像	LEDCH-S	0.7056	0.7126	0.7265	0.7389	0.7443	0.4975	0.5026	0.5142	0.5173	0.5285
	LEDCH-10K	0.7243	0.7371	0.7496	0.7488	0.7546	0.5062	0.5122	0.5203	0.5288	0.5355
	LEDCH	<b>0.8247</b>	<b>0.8395</b>	<b>0.8465</b>	<b>0.8567</b>	<b>0.8582</b>	<b>0.7718</b>	<b>0.7998</b>	<b>0.8163</b>	<b>0.8191</b>	<b>0.8230</b>

### 3.4.4 参数敏感性分析

我们也在 MIRFlickr-25K 数据集上进行了参数实验来分析 LEDCH 对参数  $\alpha$  和  $\omega$  的敏感性. 值得注意的是, 参数  $\theta$  和  $\lambda$  是避免过拟合的正则化项的系数, 通常可以根据经验将其设置为一个较小的值. 图 2 展示了 LEDCH 在图像检索文本和文本检索图像两种跨模态检索任务上的 MAP 结果. 哈希码的位数是 16. 从图中我们可以看出, 参数  $\alpha$  对准确率有一定的影响, 尤其当取值为 0 时的 MAP 值非常低, 说明本文提出的基于迁移学习的标记增强的有效性. 参数  $\theta$  对准确率也有一定的影响, 但不是非常明显, 因为它在目标函数中只是起到避免优化偏移的作用. 总之,  $\alpha$  和  $\omega$  均为 1 时, LEDCH 的表现最好.

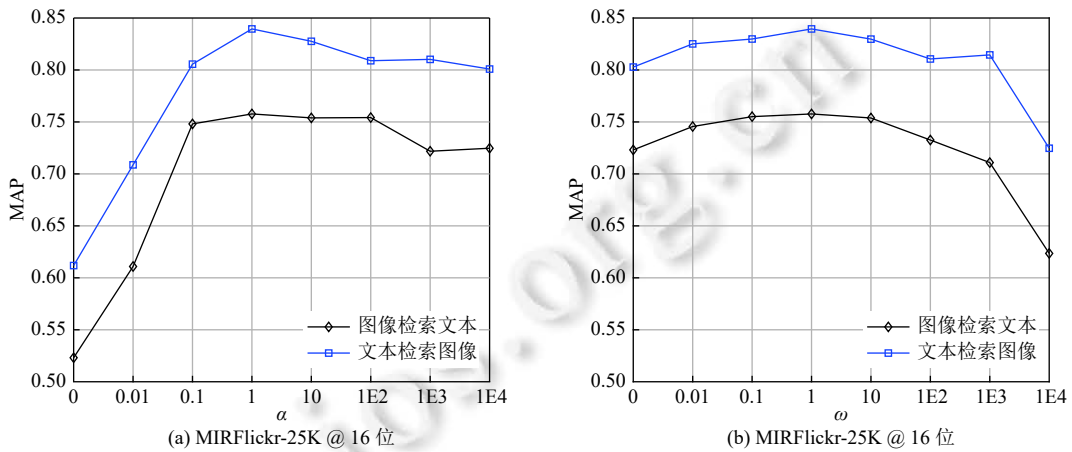


图 2 关于  $\alpha$  和  $\omega$  的参数敏感性分析

### 3.4.5 收敛性分析

本文在第 2.2.3 节和第 2.3.3 节中分别给出了标记增强和哈希码学习优化算法的收敛性证明. 为了进一步证明它们的收敛性, 图 3 给出了 LEDCH 的两个优化算法在 MIRFlickr-25K 和 NUS-WIDE 上的归一化目标损失随迭代次数变化的曲线. 哈希码的位数是 16. 从图 3 我们可以看出, 两个算法均能在 10 次迭代左右达到收敛, 证明了它的收敛性和高效性.

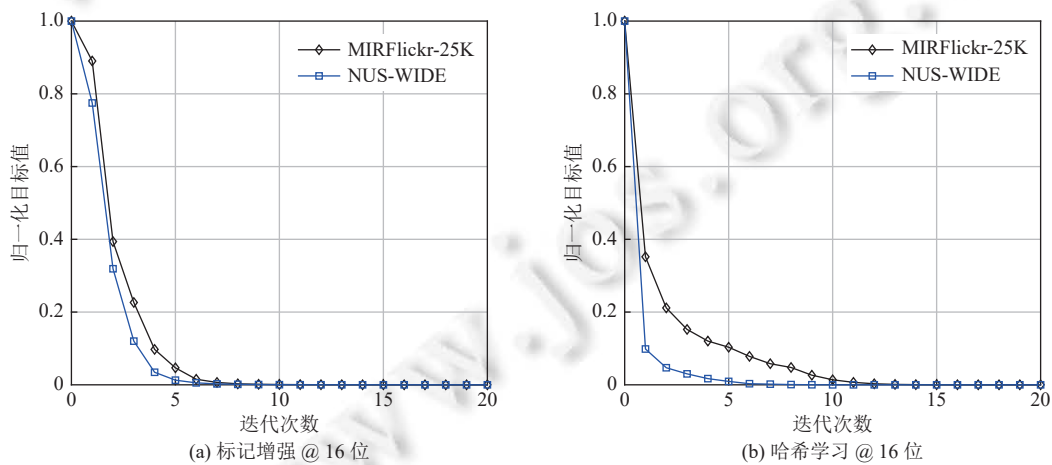


图 3 关于标记增强和哈希学习算法的收敛曲线

因为迭代次数会影响模型的准确率和效率, 标记增强优化过程是独立于哈希学习的, 它对效率相对不敏感, 可以设置一个较大的迭代次数来保证最优解. 而哈希码的离散优化算法对效率和准确率都很敏感, 需要设置一个折

中的迭代次数来同时考虑收敛率与优化效率. 因此, 在本文实验中, 两个优化算法的最大迭代次数  $t_1$  和  $t_2$  分别被设置为 10 和 5.

## 4 结 论

本文提出了一个基于标记增强的离散跨模态哈希方法 LEDCH. 它通过借鉴迁移学习引入先验知识对监督信息进行标记增强, 从而挖掘数据的真实语义分布, 并基于该分布构建相似性矩阵. 此外, 本文提出了一个高效的离散优化算法, 可以离散地求解哈希码, 减小了量化误差, 从而保证了生成哈希码的质量. 最后, 本文在两个广泛应用的基准数据集上进行了实验, 结果表明了本文方法在准确率和效率上的有效性. 将来, 我们计划将该模型扩展到在线检索场景, 并且可以通过将标记增强与协同矩阵分解结合进一步挖掘数据信息.

### References:

- [1] Wang JD, Zhang T, Song JK, Sebe N, Shen HT. A survey on learning to hash. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 769–790. [doi: [10.1109/TPAMI.2017.2699960](https://doi.org/10.1109/TPAMI.2017.2699960)]
- [2] Mao XJ, Yang YB. Semantic hashing with image subspace learning. *Ruan Jian Xue Bao/Journal of Software*, 2014, 25(8): 1781–1793 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4488.htm> [doi: [10.13328/j.cnki.jos.004488](https://doi.org/10.13328/j.cnki.jos.004488)]
- [3] Gong YC, Lazebnik S, Gordo A, Perronnin F. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013, 35(12): 2916–2929. [doi: [10.1109/TPAMI.2012.193](https://doi.org/10.1109/TPAMI.2012.193)]
- [4] Liu W, Mu C, Kumar S, Chang SF. Discrete graph hashing. In: *Proc. of the 27th Int'l Conf. on Neural Information Processing Systems*. Cambridge: MIT Press, 2014. 3419–3427.
- [5] Liu W, Wang J, Ji RR, Jiang YG, Chang SF. Supervised hashing with kernels. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Providence: IEEE, 2012. 2074–2081. [doi: [10.1109/CVPR.2012.6247912](https://doi.org/10.1109/CVPR.2012.6247912)]
- [6] Shen FM, Shen CH, Liu W, Shen HT. Supervised discrete hashing. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 37–45. [doi: [10.1109/CVPR.2015.7298598](https://doi.org/10.1109/CVPR.2015.7298598)]
- [7] Peng YX, Qi JW, Yuan YX. Modality-specific cross-modal similarity measurement with recurrent attention network. *IEEE Trans. on Image Processing*, 2018, 27(11): 5585–5599. [doi: [10.1109/TIP.2018.2852503](https://doi.org/10.1109/TIP.2018.2852503)]
- [8] Nie XS, Liu XB, Xi XM, Yin YL. Model parameter re-optimization for linear hashing based on similarity drive. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(4): 1039–1050 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5918.htm> [doi: [10.13328/j.cnki.jos.005918](https://doi.org/10.13328/j.cnki.jos.005918)]
- [9] Zhuang Y, Zhuang YT, Wu F. An integrated indexing structure for large-scale cross-media retrieval. *Ruan Jian Xue Bao/Journal of Software*, 2008, 19(10): 2667–2680 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/2667.htm> [doi: [10.3724/SP.J.1001.2008.02667](https://doi.org/10.3724/SP.J.1001.2008.02667)]
- [10] Peng YX, Qi JW. Reinforced cross-media correlation learning by context-aware bidirectional translation. *IEEE Trans. on Circuits and Systems for Video Technology*, 2020, 30(6): 1718–1731. [doi: [10.1109/TCSVT.2019.2907400](https://doi.org/10.1109/TCSVT.2019.2907400)]
- [11] Nie XS, Liu XB, Xi XM, Li CL, Yin YL. Fast unmediated hashing for cross-modal retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 2021, 31(9): 3669–3678. [doi: [10.1109/TCSVT.2020.3042972](https://doi.org/10.1109/TCSVT.2020.3042972)]
- [12] Deng C, Chen ZJ, Liu XL, Gao XB, Tao DC. Triplet-based deep hashing network for cross-modal retrieval. *IEEE Trans. on Image Processing*, 2018, 27(8): 3893–3903. [doi: [10.1109/TIP.2018.2821921](https://doi.org/10.1109/TIP.2018.2821921)]
- [13] Peng YX, Qi JW, Huang X, Yuan YX. CCL: Cross-modal correlation learning with multigrained fusion by hierarchical network. *IEEE Trans. on Multimedia*, 2017, 20(2): 405–420. [doi: [10.1109/TMM.2017.2742704](https://doi.org/10.1109/TMM.2017.2742704)]
- [14] Liu H, Lin MB, Zhang SC, Wu YL, Huang FY, Ji RR. Dense auto-encoder hashing for robust cross-modality retrieval. In: *Proc. of the 26th ACM Int'l Conf. on Multimedia*. Seoul: Association for Computing Machinery, 2018. 1589–1597. [doi: [10.1145/3240508.3240684](https://doi.org/10.1145/3240508.3240684)]
- [15] Wang YX, Chen ZD, Luo X, Li R, Xu XS. Fast cross-modal hashing with global and local similarity embedding. *IEEE Trans. on Cybernetics*, 2021: 1–14. [doi: [10.1109/TCYB.2021.3059886](https://doi.org/10.1109/TCYB.2021.3059886)]
- [16] Ding GG, Guo YC, Zhou JL. Collective matrix factorization hashing for multimodal data. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 2083–2090. [doi: [10.1109/CVPR.2014.267](https://doi.org/10.1109/CVPR.2014.267)]
- [17] Liu H, Ji RR, Wu YJ, Huang FY, Zhang BC. Cross-modality binary code learning via fusion similarity hashing. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 6345–6353. [doi: [10.1109/CVPR.2017.672](https://doi.org/10.1109/CVPR.2017.672)]
- [18] Wang D, Wang Q, Gao XB. Robust and flexible discrete hashing for cross-modal similarity search. *IEEE Trans. on Circuits and Systems*

- for Video Technology, 2018, 28(10): 2703–2715. [doi: [10.1109/TCSVT.2017.2723302](https://doi.org/10.1109/TCSVT.2017.2723302)]
- [19] Hu MQ, Yang Y, Shen FM, Xie N, Hong RC, Shen HT. Collective reconstructive embeddings for cross-modal hashing. *IEEE Trans. on Image Processing*, 2019, 28(6): 2770–2784. [doi: [10.1109/TIP.2018.2890144](https://doi.org/10.1109/TIP.2018.2890144)]
- [20] Lin ZJ, Ding GG, Hu MQ, Wang JM. Semantics-preserving hashing for cross-view retrieval. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 3864–3872. [doi: [10.1109/CVPR.2015.7299011](https://doi.org/10.1109/CVPR.2015.7299011)]
- [21] Zhang PF, Li CX, Liu MY, Nie LQ, Xu XS. Semi-relaxation supervised hashing for cross-modal retrieval. In: *Proc. of the 25th ACM Int'l Conf. on Multimedia*. Mountain View: Association for Computing Machinery, 2017. 1762–1770. [doi: [10.1145/3123266.3123320](https://doi.org/10.1145/3123266.3123320)]
- [22] Mandal D, Chaudhury KN, Biswas S. Generalized semantic preserving hashing for cross-modal retrieval. *IEEE Trans. on Image Processing*, 2019, 28(1): 102–112. [doi: [10.1109/TIP.2018.2863040](https://doi.org/10.1109/TIP.2018.2863040)]
- [23] Zhang DQ, Li WJ. Large-scale supervised multimodal hashing with semantic correlation maximization. In: *Proc. of the 28th AAAI Conf. on Artificial Intelligence*. Quebec City: AAAI Press, 2014. 2177–2183.
- [24] Xu X, Shen FM, Yang Y, Shen HT, Li XL. Learning discriminative binary codes for large-scale cross-modal retrieval. *IEEE Trans. on Image Processing*, 2017, 26(5): 2494–2507. [doi: [10.1109/TIP.2017.2676345](https://doi.org/10.1109/TIP.2017.2676345)]
- [25] Chen ZD, Li CX, Luo X, Nie LQ, Zhang W, Xu XS. SCRATCH: A scalable discrete matrix factorization hashing framework for cross-modal retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 2020, 30(7): 2262–2275. [doi: [10.1109/TCSVT.2019.2911359](https://doi.org/10.1109/TCSVT.2019.2911359)]
- [26] Jiang QY, Li WJ. Discrete latent factor model for cross-modal hashing. *IEEE Trans. on Image Processing*, 2019, 28(7): 3490–3501. [doi: [10.1109/TIP.2019.2897944](https://doi.org/10.1109/TIP.2019.2897944)]
- [27] Kong J, Sun QS, Xu H, Liu YZ, Ji ZX. Multi-object classification of remote sensing image based on affine-invariant supervised discrete hashing. *Ruan Jian Xue Bao/Journal of Software*, 2019, 30(4): 914–926 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5661.htm> [doi: [10.13328/j.cnki.jos.005661](https://doi.org/10.13328/j.cnki.jos.005661)]
- [28] Wang D, Gao XB, Wang XM, He LH. Label consistent matrix factorization hashing for large-scale cross-modal similarity search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, 41(10): 2466–2479. [doi: [10.1109/TPAMI.2018.2861000](https://doi.org/10.1109/TPAMI.2018.2861000)]
- [29] Gui J, Liu TL, Sun ZN, Tao DC, Tan TN. Supervised discrete hashing with relaxation. *IEEE Trans. on Neural Networks and Learning Systems*, 2016, 29(3): 608–617. [doi: [10.1109/TNNLS.2016.2636870](https://doi.org/10.1109/TNNLS.2016.2636870)]
- [30] Chen ZD, Yu WJ, Li CX, Nie LQ, Xu XS. Dual deep neural networks cross-modal hashing. In: *Proc. of the 32nd AAAI Conf. on Artificial Intelligence*. New Orleans: AAAI Press, 2018. 274–281.
- [31] Jiang QY, Li WJ. Deep cross-modal hashing. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 3270–3278. [doi: [10.1109/CVPR.2017.348](https://doi.org/10.1109/CVPR.2017.348)]
- [32] Jin S, Zhou SC, Liu Y, Chen C, Sun XS, Yao HX, Hua XS. SSAH: Semi-supervised adversarial deep hashing with self-paced hard sample generation. In: *Proc. of the 34th AAAI Conf. on Artificial Intelligence*. New York: AAAI Press, 2020. 11157–11164. [doi: [10.1609/aaai.v34i07.6773](https://doi.org/10.1609/aaai.v34i07.6773)]
- [33] Bai C, Zeng C, Ma Q, Zhang JL, Chen SY. Deep adversarial discrete hashing for cross-modal retrieval. In: *Proc. of the 2020 Int'l Conf. on Multimedia Retrieval*. Dublin: Association for Computing Machinery, 2020. 525–531. [doi: [10.1145/3372278.3390711](https://doi.org/10.1145/3372278.3390711)]
- [34] Wang YX, Luo X, Xu XS. Label embedding online hashing for cross-modal retrieval. In: *Proc. of the 28th ACM Int'l Conf. on Multimedia*. Seattle: Association for Computing Machinery, 2020. 871–879. [doi: [10.1145/3394171.3413971](https://doi.org/10.1145/3394171.3413971)]
- [35] Xu N, Tao AP, Geng X. Label enhancement for label distribution learning. In: *Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence*. Stockholm: IJCAI.org, 2018. 2926–2932. [doi: [10.24963/ijcai.2018/406](https://doi.org/10.24963/ijcai.2018/406)]
- [36] Xu N, Shu J, Liu YP, Geng X. Variational label enhancement. In: *Proc. of the 37th Int'l Conf. on Machine Learning*. Virtual Event: PMLR, 2020. 10597–10606.
- [37] Yang Y, Luo YD, Chen WL, Shen FM, Shao J, Shen HT. Zero-shot hashing via transferring supervised knowledge. In: *Proc. of the 24th ACM Int'l Conf. on Multimedia*. Amsterdam: Association for Computing Machinery, 2016. 1286–1295. [doi: [10.1145/2964284.2964319](https://doi.org/10.1145/2964284.2964319)]
- [38] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: *Proc. of the 26th Int'l Conf. on Neural Information Processing Systems*. Red Hook: Curran Associates Inc., 2013. 3111–3119.
- [39] Rudin W. *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976. 3.
- [40] Chen ZD, Wang YX, Li HQ, Luo X, Nie LQ, Xu XS. A two-step cross-modal hashing by exploiting label correlations and preserving similarity in both steps. In: *Proc. of the 27th ACM Int'l Conf. on Multimedia*. Nice: Association for Computing Machinery, 2019. 1694–1702. [doi: [10.1145/3343031.3350862](https://doi.org/10.1145/3343031.3350862)]
- [41] Wang YX, Luo X, Nie LQ, Song JK, Zhang W, Xu XS. BATCH: A scalable asymmetric discrete cross-modal hashing. *IEEE Trans. on Knowledge and Data Engineering*, 2021, 33(11): 3507–3519. [doi: [10.1109/TKDE.2020.2974825](https://doi.org/10.1109/TKDE.2020.2974825)]

- [42] Li CX. Research on discrete hashing method based on matrix factorization [MS. Thesis]. Jinan: Shandong University, 2019 (in Chinese with English abstract).
- [43] Luo X. Large-scale media retrieval based on learning to hash [Ph.D. Thesis]. Jinan: Shandong University, 2019 (in Chinese with English abstract).
- [44] Huijskes MJ, Lew MS. The MIR Flickr retrieval evaluation. In: Proc. of the 1st ACM Int'l Conf. on Multimedia Information Retrieval. Vancouver: Association for Computing Machinery, 2008. 39–43. [doi: 10.1145/1460096.1460104]
- [45] Chua TS, Tang JH, Hong RC, Li HJ, Luo ZP, Zheng YT. NUS-WIDE: A real-world Web image database from National University of Singapore. In: Proc. of the 8th ACM Int'l Conf. on Image and Video Retrieval. Santorini Island: Association for Computing Machinery, 2009. 48. [doi: 10.1145/1646396.1646452]
- [46] Liu XB, Nie XS, Zeng WJ, Cui CR, Zhu L, Yin YL. Fast discrete cross-modal hashing with regressing from semantic labels. In: Proc. of the 26th ACM Int'l Conf. on Multimedia. Seoul: Association for Computing Machinery, 2019. 1662–1669. [doi: 10.1145/3240508.3240683]
- [47] Shen HT, Liu LC, Yang Y, Xu X, Huang Z, Shen FM, Hong RC. Exploiting subspace relation in semantic labels for cross-modal hashing. IEEE Trans. on Knowledge and Data Engineering, 2021, 33(10): 3351–3365. [doi: 10.1109/TKDE.2020.2970050]

#### 附中文参考文献:

- [2] 毛晓蛟, 杨育彬. 一种基于子空间学习的图像语义哈希索引方法. 软件学报, 2014, 25(8): 1781–1793. <http://www.jos.org.cn/1000-9825/4488.htm> [doi: 10.13328/j.cnki.jos.004488]
- [8] 聂秀山, 刘兴波, 裘肖明, 尹义龙. 基于相似度驱动的线性哈希模型参数再优化方法. 软件学报, 2020, 31(4): 1039–1050. <http://www.jos.org.cn/1000-9825/5918.htm> [doi: 10.13328/j.cnki.jos.005918]
- [9] 庄毅, 庄越挺, 吴飞. 一种支持海量跨媒体检索的集成索引结构. 软件学报, 2008, 19(10): 2667–2680. <http://www.jos.org.cn/1000-9825/19/2667.htm> [doi: 10.3724/SP.J.1001.2008.02667]
- [27] 孔颀, 孙权森, 徐晖, 刘亚洲, 纪则轩. 基于仿射不变离散哈希的遥感图像多目标分类. 软件学报, 2019, 30(4): 914–926. <http://www.jos.org.cn/1000-9825/5661.htm> [doi: 10.13328/j.cnki.jos.005661]
- [42] 李传祥. 基于矩阵分解的离散哈希方法研究 [硕士学位论文]. 济南: 山东大学, 2019.
- [43] 罗昕. 基于哈希学习的大规模媒体检索研究 [博士学位论文]. 济南: 山东大学, 2019.



王永欣(1992—), 女, 博士, 副教授, 主要研究领域为信息检索, 机器学习, 计算机视觉.



罗昕(1992—), 男, 博士, 助理教授, CCF 专业会员, 主要研究领域为机器学习, 多媒体检索, 计算机视觉.



田洁茹(1997—), 女, 硕士, 主要研究领域为计算机视觉, 机器学习, 医学图像分析.



许信顺(1975—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习, 信息检索, 数据挖掘, 图像/视频分析与检索.



陈振铎(1993—), 男, 博士, CCF 专业会员, 主要研究领域为机器学习, 信息检索, 图像/视频分析与检索.