

面向语义增强的在线哈希方法

赵志杰¹, 康 潇², 张雪凝², 王少华¹, 刘兴波¹, 聂秀山¹

(1. 山东建筑大学计算机与人工智能学院, 济南 250101; 2. 山东大学软件学院, 济南 250101)

摘要: 传统的基于批处理的哈希学习方法通常无法满足大规模流数据实时在线检索的需求。在线哈希学习其核心在于无需重复访问原始累积数据, 只为新增数据学习哈希码, 并实时更新哈希函数以适应新旧数据的变化。现有在线哈希方法仍面临诸多挑战, 如类间关系挖掘不足导致的语义偏移和新旧数据关联不足导致的遗忘问题。针对这些问题, 本文提出了一种新的在线哈希学习方法——面向语义增强的在线哈希(Online semantic enhancement hashing, OSEH)。该方法通过设计三重矩阵分解框架, 深入挖掘特征和标签间的交互关系, 以生成反映类间关系的细粒度标签矩阵。同时, 结合标签嵌入和成对相似性保持技术, 将增强的语义信息有效融入哈希学习过程, 优化哈希码的生成和哈希函数的实时更新。在大规模检索数据集上的实验结果表明, 本文所提方法显著提升了在线哈希学习的性能。

关键词: 在线检索; 监督在线哈希; 三重矩阵分解; 细粒度标签

中图分类号: TP391 **文献标志码:** A

Online Semantic Enhancement Hashing

ZHAO Zhijie¹, KANG Xiao², ZHANG Xuening², WANG Shaohua¹, LIU Xingbo¹, NIE Xiushan¹

(1. School of Computer and Artificial Intelligence, Shandong Jianzhu University, Jinan 250101, China; 2. School of Software, Shandong University, Jinan 250101, China)

Abstract: Batch-based hash learning methods are usually inadequate for real-time online retrieval of large-scale streaming data. Therefore, online hashing has emerged as a promising solution, enabling the learning of hash codes for new data without revisiting old data and adapting hash functions to coming data. However, several challenges persist, including semantic drift caused by insufficient exploration of inter-class relationships and data forgetting resulting from limited association between new and old data. To address these challenges, this paper proposes a novel supervised method named online semantic enhancement hashing (OSEH). It designs a triple matrix factorization framework, which mutually bridges the gap of original features and one-hot labels, thereafter constructing a fine-grained label matrix. Moreover, by seamlessly integrating label embedding and pairwise similarity, the proposed method effectively embeds enriched semantics into the process of hash learning, optimizing both hash code and function. Experimental evaluations conducted on benchmark datasets validate the effectiveness of the proposed method.

Key words: online retrieval; supervised online hashing; tri-matrix factorization; fine-grained labels

引 言

随着互联网中媒体数据的急剧增长,大规模数据检索面临着诸多挑战。哈希学习作为近似最近邻检索的典型方法,凭借其卓越的存储效率和查询速度,受到了研究者的广泛关注^[1-2]。然而,传统的哈希学习算法通常在一个封闭且静态的环境中进行训练,即用预先收集完整的数据集训练出相应的哈希码和哈希函数。然而在现实应用中,数据往往并非一次性收集齐全,而是以流数据的形式持续、动态地涌入。例如,在社交媒体平台上,用户时刻都在上传新的图文内容;在推荐系统中,用户的行为和偏好也在不断变化。在这种情境下,传统的基于批处理的哈希算法每当需要更新模型时,都不得不重新访问和处理原始累积数据,进而带来昂贵的更新成本^[3-4]。为了解决这个问题,近年来研究者们提出了在线哈希学习的研究思路,其旨在学习新增数据的哈希码,更新哈希函数,同时确保其对全体数据的适用性,从而实现实时在线检索。

根据是否利用了监督信息,现有的在线哈希方法可以大致划分为无监督在线哈希和监督在线哈希。无监督在线哈希方法通常采用基于在线优化或者数据草图的方案来平衡对累积数据的记忆和对新增数据的适应^[5-6]。监督在线哈希方法则主要挖掘流数据间一致性的语义标签来指导哈希码的生成和更新^[7-10]。得益于监督信息的指导,监督在线哈希方法通常较无监督方法有显著的性能提升。然而,现有的监督在线哈希方法仍存在问题有待进一步探索和研究。首先,这些方法通常采用独热码标签作为监督信息,这类标签不同类别之间海明距离统一固定为2,无法表征细粒度的类间相似关系。例如,狗和狼之间的距离应该比狗和鸟之间的距离更近,但独热码无法反映这种关系。大量研究表明,类间关系对于缓解在线哈希学习中的语义偏移问题至关重要^[11-12]。其次,许多现有的方法仅仅依赖标签信息来指导哈希码的生成,忽略了对新旧数据间成对相似关系的保持,导致累积数据记忆不充分,从而影响检索性能。

为了解决上述问题,本文提出一种面向语义增强的在线哈希(Online semantic enhancement hashing, OSEH)学习方法。该方法设计一种基于三重矩阵分解的学习框架来构建特征空间和语义标签之间的关联关系。通过特征和标签之间的交互映射来学习样本属性信息,进而生成细粒度语义标签。该方法整合正交约束、离散约束和非对称成对相似关系保持,将所学的细粒度语义标签嵌入到哈希学习过程中,指导哈希码的生成和哈希函数的在线更新。

本文的主要贡献包括以下两个方面:(1)提出了在线哈希学习中一种新的监督信息利用方法,通过特征和标签之间的互补学习实现语义增强,有效融合特征空间中的样本相似关系与独热码标签的类别区分信息,生成蕴含更丰富语义的细粒度标签,从而缓解流数据语义偏移问题。(2)设计了统一的学习与优化框架,通过整合多重约束条件,将所学的细粒度语义标签和非对称成对相似关系嵌入到哈希学习过程中,从而平衡对累积数据的记忆和对新增数据的适应。

1 相关工作

1.1 无监督哈希方法

无监督在线哈希表示学习过程不需要利用训练样本的标记信息,这种方法可通过捕捉数据之间的相似性,并将相似的数据映射到相邻的二值空间位置,以实现高效的数据检索和相似性匹配^[13]。无监督在线哈希方法的优势在于不需要额外的标注工作,可以适用于大规模数据集。在无监督方法中,在线核哈希(Online kernel hashing, OKH)^[14]提出了一种新的零均值策略,引入核函数,实现了在线优化。但是此方法一次只能处理一对样本,总体实施起来效率较低。基于多位置量化的无监督在线哈希(Unsu-

pervised online hashing with multi-bit quantization, OHMBQ)^[15]提出了一种新的多位量化的无监督在线哈希方法,通过维护小尺寸的数据草图来保存流数据信息,从数据草图中学习正交变换,保证流数据内部分量相互独立。然后学习最优量化器,通过建立数据分布模型,自适应地用多比特量化每个分量,以更好地保持数据的在线相似性。

1.2 监督哈希方法

监督在线哈希方法主要依靠标签信息或成对相似关系来学习哈希码和哈希函数,通过使用标记数据集中的类别或相似性信息,监督在线哈希方法可以更准确地学习到数据之间的相似性关系,并生成具有区分性的哈希码。监督在线哈希(Online supervised hashing, OSH)^[16]基于 ECOC(Error correct output code)编码向量,不设置类标签的数量,通过优化与哈希码之间的损失来学习哈希函数。基于哈达玛矩阵的在线哈希(Hadamard matrix guided online hashing, HMOH)^[17]通过引入哈达玛矩阵来指导哈希函数的学习,大大提高了训练效率。平衡相似度在线离散哈希(Balance similarity online discrete hashing, BSODH)^[18]设计了一个平衡相似度矩阵来处理数据不平衡问题,但是此方法在线自适应性较差,待优化变量多。同样是依赖相似度矩阵学习,基于多标记投影的快速在线哈希算法(Fast online hashing with multi-label projection, FOH)^[19]只更新数据库中中小部分的二值码,并考虑多标签信息,构造相似度矩阵,通过引入多标签投影损失,以此保证数据之间的相似性。基于标签语义增强的鲁棒在线哈希算法(Robust online hashing with label semantic enhancement, ROHLSE)^[20]通过将噪声标签恢复干净标签,并根据样本和标签之间的关系对标签进行预测,充分利用了标签之间的语义相关性。基于哈希中心的自蒸馏双存储在线哈希方法(Self-distillation dual-memory online hashing with hash centers, SDOH-HC)^[21]设计了重放和蒸馏两个模块。对于重放模块,本文提出了一种包含哈希中心和样本的双存储机制,该模块可以更好地学习和利用历史信息。对于知识蒸馏模块,可以使哈希中心从自身提取信息,所提取的信息来自于上一轮的版本。

2 OSEH 算法

2.1 符号定义

现给定训练样本集 X ,前 $t-1$ 轮累计数据用 $X^{(t-1)}$ 表示,在进行第 t 轮的训练时,新到达的数据用 $X^{(t)}$ 表示,其中, $X^{(t-1)} \in \mathbf{R}^{d \times N_t}$, $X^{(t)} \in \mathbf{R}^{d \times n_t}$, d 表示样本维度, n_t 为新到达的样本个数, N_t 为上一轮累计的样本个数。哈希码由 $B = [B^{(t)}, B^{(t-1)}]$ 组成,其中 $B^{(t)} \in \{-1, 1\}^{r \times n_t}$ 和 $B^{(t-1)} \in \{-1, 1\}^{r \times N_t}$ 分别对应新旧数据, r 为哈希码长。 $L^{(t)} \in \mathbf{R}^{c \times n_t}$ 和 $L^{(t-1)} \in \mathbf{R}^{c \times N_t}$ 为样本的标签, c 代表类别数。

2.2 建模过程

本文所提的 OSEH 方法的框架图如图 1 所示,该方法主要包括特征矩阵分解、细粒度标签嵌入以及相似性保持 3 个模块。

2.2.1 特征矩阵分解

由于独热码标签不同类别之间海明距离没有足够的区分性,无法表征细粒度的类间关系,为此从矩阵分解的角度考虑,将特征中所反映的细粒度样本关系和独热码反映的类别区分信息结合在一起,之后提取出属性,用所得属性辅助细粒度标签的生成,以学习更具区分性的哈希码。具体来说,将特征矩阵 X 分解为正交分类映射、属性矩阵和标签 3 部分, $A \in \mathbf{R}^{k \times c}$ 为属性矩阵,表示在每个类别中含有不同的属性, $D \in \mathbf{R}^{d \times k}$ 为正交分类映射, L 为标签。这个过程旨在完成属性矩阵到特征之间的映射,另外

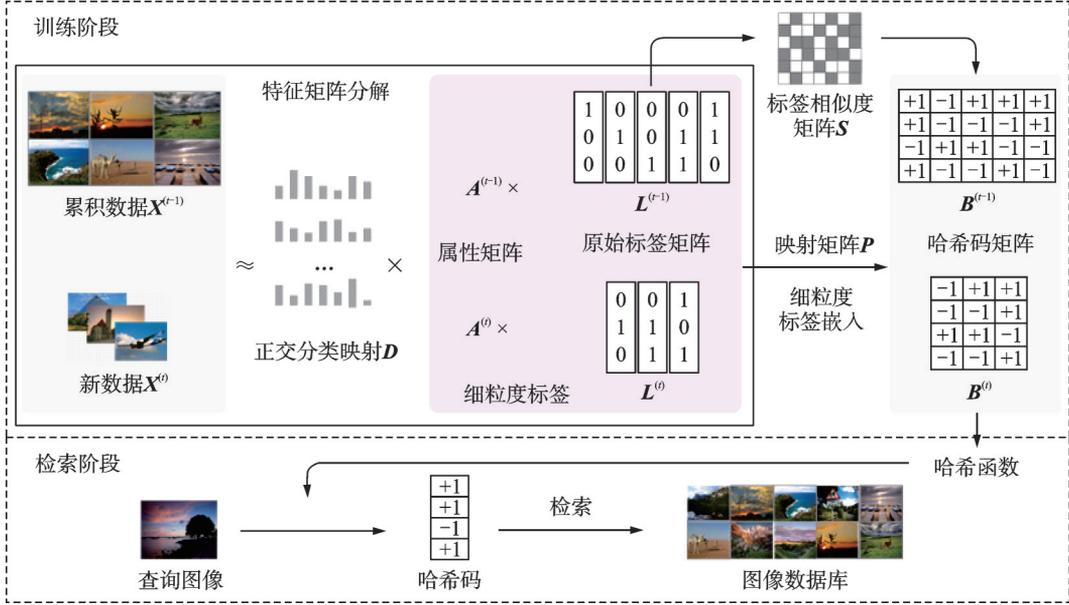


图1 OSEH方法框架图

Fig.1 Framework of the proposed OSEH

对 D 和 A 设置正交约束以保证特征和类别之间能够相互独立,去除冗余信息。这个过程可以形式化表述为

$$\begin{cases} \min_{D, A^{(t)}} \|X^{(t)} - DA^{(t)}L^{(t)}\|_F^2 + \|X^{(t-1)} - DA^{(t-1)}L^{(t-1)}\|_F^2 \\ \text{s.t. } D^T D = I_k, A^{(t)T} A^{(t)} = I_c \end{cases} \quad (1)$$

2.2.2 细粒度标签嵌入

由于独热码标签无法表征细粒度的类间关系,本文提出使用细粒度标签嵌入与特征标签联合学习的策略促进语义信息的挖掘,用属性矩阵和标签结合表示新学到的细粒度标签,以表示出样本间的更细微差别。 $P \in \mathbb{R}^{r \times k}$ 为映射矩阵,将学到的细粒度标签经过 P 映射到海明空间中,以此更好地指导学习哈希码。这个过程可以形式化表述为

$$\begin{cases} \min_{B^{(t)}, B^{(t-1)}, P, A^{(t)}} \|B^{(t)} - PA^{(t)}L^{(t)}\|_F^2 + \|B^{(t-1)} - PA^{(t-1)}L^{(t-1)}\|_F^2 \\ \text{s.t. } B^{(t)} \in \{-1, +1\}^{r \times n_t}, B^{(t-1)} \in \{-1, +1\}^{r \times N_t}, A^{(t)T} A^{(t)} = I_c \end{cases} \quad (2)$$

2.2.3 相似性保持

为了充分考虑新旧数据之间的关系,选取非对称相似度保持策略,采用最小化内积和相似度矩阵间的距离来保证和两两样本之间的相似关系,利用相似关系训练哈希码,这个过程可以形式化表述为

$$\begin{cases} \min_{B^{(t)}, B^{(t-1)}} \|B^{(t)T} B^{(t-1)} - rS\|_F^2 \\ \text{s.t. } B^{(t)} \in \{-1, +1\}^{r \times n_t}, B^{(t-1)} \in \{-1, +1\}^{r \times N_t} \end{cases} \quad (3)$$

式中: $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数; $S \in \{-1, 1\}^{N_t \times n_t}$ 为标签相似度矩阵,在此模块约束新旧哈希码之间的相似关系保持,当样本之间共享至少一个标签时 $S_{i,j} = 1$, 否则 $S_{i,j} = -1$ 。

2.2.4 目标函数

结合式(1~3)最终的目标函数设计如下

$$\begin{cases} \min_{A^{(t)}, P, B^{(t)}, B^{(t-1)}, D, W} \alpha \left(\|X^{(t)} - DA^{(t)}L^{(t)}\|_F^2 + \|X^{(t-1)} - DA^{(t-1)}L^{(t-1)}\|_F^2 \right) + \|B^{(t)T}B^{(t-1)} - rS\|_F^2 + \\ \sigma \|W^T X^{(t)} - B^{(t)}\|_F^2 + \theta \left(\|B^{(t)} - PA^{(t)}L^{(t)}\|_F^2 + \|B^{(t-1)} - PA^{(t-1)}L^{(t-1)}\|_F^2 \right) + \lambda \|W\|_F^2 + \gamma \|P\|_F^2 \\ \text{s.t. } D^T D = I_k, A^{(t)T} A^{(t)} = I_r, B^{(t)} \in \{-1, +1\}^{r \times n_t}, B^{(t-1)} \in \{-1, +1\}^{r \times N_t} \end{cases} \quad (4)$$

式中: W 为映射矩阵; α, σ, θ 为平衡参数; λ, γ 为正则化参数。

2.3 优化求解

式(4)中定义的目标函数是一个非凸非连续问题,难以直接优化。因此设计一个迭代优化算法,分别求解各个变量相关的子问题,然后通过变量之间的迭代优化来试图找到问题的局部最优解。

(1) $B^{(t)}$ -step: 固定除 $B^{(t)}$ 以外的其他变量,式(4)中的问题可以重写为

$$\begin{cases} \min_{B^{(t)}} \|B^{(t)T}B^{(t-1)} - rS\|_F^2 + \sigma \|W^T X^{(t)} - B^{(t)}\|_F^2 + \theta \|B^{(t)} - PA^{(t)}L^{(t)}\|_F^2 \\ \text{s.t. } B^{(t)} \in \{-1, +1\}^{r \times n_t} \end{cases} \quad (5)$$

等价于

$$\min_{B^{(t)}} \|B^{(t-1)T}B^{(t)}\|_F^2 - 2\text{tr}(H^T B^{(t)}) \quad (6)$$

式中 $H = kB^{(t-1)}S^T + \sigma W^T X^{(t)} + \theta PA^{(t)}L^{(t)}$ 。由于式(6)中 $B^{(t)}$ 的求解是 NP-hard 问题,难以直接求解,于是使用离散循环坐标下降法求解该子问题。将哈希码逐位求解,定义 $\tilde{b}_i^{(t)}, \tilde{b}_i^{(t-1)}$ 分别是 $B^{(t)}$ 和 $B^{(t-1)}$ 第 i 行要更新的位, \tilde{h}_i 则代表 H 的第 i 行。 $\tilde{B}^{(t)}, \tilde{B}^{(t-1)}, \tilde{H}$ 分别是矩阵 $B^{(t)}, B^{(t-1)}, H$ 去掉 $\tilde{b}_i^{(t)}, \tilde{b}_i^{(t-1)}, \tilde{h}_i$ 的形式,上述问题等价于

$$\begin{cases} \min_{\tilde{b}_i^{(t)}} \text{tr} \left((\tilde{B}^{(t)T} \tilde{B}^{(t-1)} \tilde{b}_i^{(t-1)} - \tilde{h}_i^T) \tilde{b}_i^{(t)} \right) \\ \text{s.t. } \tilde{b}_i^{(t)} \in \{-1, +1\}^{r \times n_t} \end{cases} \quad (7)$$

更新规则如下

$$\tilde{b}_i^{(t)} = \text{sgn}(\tilde{h}_i - \tilde{b}_i^{(t-1)} \tilde{B}^{(t-1)T} \tilde{B}^{(t)}) \quad (8)$$

$B^{(t-1)}$ -step: 固定除 $B^{(t-1)}$ 以外的其他变量,式(4)中的问题可以重写为

$$\begin{cases} \min_{B^{(t-1)}} \|B^{(t)T}B^{(t-1)} - rS\|_F^2 + \theta \|B^{(t-1)} - PA^{(t-1)}L^{(t-1)}\|_F^2 \\ \text{s.t. } B^{(t-1)} \in \{-1, +1\}^{r \times N_t} \end{cases} \quad (9)$$

等价于

$$\min_{B^{(t-1)}} \text{tr} \left(B^{(t-1)T} (B^{(t)} B^{(t)T} B^{(t-1)} - 2Z) \right) \quad (10)$$

式中 $Z = rB^{(t)}S + \theta PA^{(t-1)}L^{(t-1)}$,式(9)的最优解如下

$$B^{(t-1)} = \text{sgn}(2Z - B^{(t)} B^{(t)T} B^{(t-1)}) \quad (11)$$

式中 $\text{sgn}(\cdot)$ 是符号函数。

(2) P -step: 固定除 P 以外的其他变量,式(4)中的问题可以重写为

$$\min_P \theta \left(\|B^{(t)} - PA^{(t)}L^{(t)}\|_F^2 + \|B^{(t-1)} - PA^{(t-1)}L^{(t-1)}\|_F^2 \right) + \gamma \|P\|_F^2 \quad (12)$$

式(12)定义了一个最小二乘问题,可以直接使用最小二乘法求解,令其导数为0,所得 P 的最优解为

$$P = \theta (B^{(t)} L^{(t)T} A^{(t)T} + B^{(t-1)} L^{(t-1)T} A^{(t-1)T}) (\theta E + \gamma I_k)^{-1} \quad (13)$$

式中 $E = A^{(t)} L^{(t)} L^{(t)T} A^{(t)T} + A^{(t-1)} L^{(t-1)} L^{(t-1)T} A^{(t-1)T}$ 。

(3) D -step: 固定除 D 以外的其他变量,式(4)中的问题可以重写为

$$\begin{cases} \min_D \alpha \left(\|X^{(t)} - DA^{(t)} L^{(t)}\|_F^2 + \|X^{(t-1)} - DA^{(t-1)} L^{(t-1)}\|_F^2 \right) \\ \text{s.t. } D^T D = I_k \end{cases} \quad (14)$$

对于式(14),通过使用奇异值分解求得此变量的最优解。令 $F_1 \Sigma O_1^T = X^{(t)} L^{(t)T} A^{(t)T} + X^{(t-1)} L^{(t-1)T} A^{(t-1)T}$,由此可以得出最优解为

$$D = F_1 \hat{O}_1^T \quad (15)$$

(4) A -step: 固定除 $A^{(t)}$ 以外的其他变量,式(4)中的问题可以重写为

$$\begin{cases} \min_{A^{(t)}} \alpha \|X^{(t)} - DA^{(t)} L^{(t)}\|_F^2 + \theta \|B^{(t)} - PA^{(t)} L^{(t)}\|_F^2 \\ \text{s.t. } A^{(t)T} A^{(t)} = I_c \end{cases} \quad (16)$$

与求解变量 D 的过程类似,可以通过奇异值分解,令 $F_2 \Sigma O_2^T = \alpha D^T X^{(t)} L^{(t)T} + \theta P^T B^{(t)} L^{(t)T}$,由此可以得出最优解为

$$A^{(t)} = F_2 \hat{O}_2^T \quad (17)$$

(5) W -step: 对于映射矩阵 W 的更新,固定其他无关变量,将 W 的导数设为0,最终得出 W 的最优解为

$$W^T = \sigma \left(\sigma (X^{(t)} X^{(t)T} + \lambda I_d)^{-1} X^{(t)} B^{(t)T} \right) \quad (18)$$

为清楚起见,在算法1中对优化的迭代过程进行了总结。

算法1 第 t 轮次的OSEH算法流程

输入:流数据特征表示 X ,第 t 轮 $X^{(t)}$ 、 $L^{(t)}$,已经存在的 $X^{(t-1)}$ 、 $L^{(t-1)}$,参数 $\alpha, \sigma, \theta, \lambda, \gamma$; c 超参数 k ,哈希码长 r , T 为迭代次数。

输出:哈希码 B ,映射矩阵 W 。

for $t = 1 : T$ do:

$B^{(t)}$ -step: 通过式(8)求解 $B^{(t)}$

$B^{(t-1)}$ -step: 通过式(11)求解 $B^{(t-1)}$

P -step: 通过式(13)求解 P

D -step: 通过式(15)求解 D

A -step: 通过式(17)求解 $A^{(t)}$

End for

设置 $B^{(t-1)} = [B^{(t)}, B^{(t-1)}]$

返回 哈希码 $B^{(t)}$,映射矩阵 W 。

3 实验及分析

本节首先简单介绍所用数据集以及相应的实验设置,之后进行实验结果的具体分析。实验选择在MIRFLICKR^[22]与NUSWIDE^[23]这2个广泛使用的数据集上来验证该方法的有效性。

3.1 数据集

(1) MIRFLICKR。该数据集共包含 20 015 张图片,使用的标签共 24 个类。选取 512 维 GIST (Generalized search trees)特征描述该数据集中图片模态,从中随机选取 2 000 个样本对作为测试集,其余的作为训练集。

(2) NUSWIDE。该数据集共包含 184 577 张图片,本文使用样本数目共有 10 个类别,使用 500 维 SIFT(Scale-invariant feature transform)特征描述该数据集中图片模态,随机选取 2 000 个样本对作为测试集,为了节省时间选取其中的 40 000 个样本作为训练集。

3.2 实验设置

本文的评估标准选择用平均精确率均值(mAP@100)和前 100 精度(Precision@100)这 2 个常用指标,对于选取的对比方法则使用作者提供的参数。对于本文所提方法,在 MIRFLICKR 数据集中将参数设置如下: $\alpha = 0.5, \sigma = 0.2, \theta = 0.1$,正则化参数 $\lambda = 0.6, \gamma = 5$;在 NUSWIDE 数据集中 $\alpha = 0.5, \sigma = 0.2, \theta = 0.1$,正则化参数 $\lambda = 0.6, \gamma = 0.01$,属性维度 k 在这 2 个数据集中都设置 100。此外,根据先前经验,将迭代次数 T 设置为 5。所有的实验结果均在 Intel(R) Core(TM) i7-11700 @2.50 GHz、32.0 GB RAM 的台式机上运行得到。

3.3 实验结果及分析

3.3.1 有效性实验

为了验证本文方法的有效性,本文所提的方法与当前流行的在线哈希方法模型 BSODH^[18]、HCOH^[24]、HMOH^[17]和 FOH^[19]进行比较,在对比试验中,FOH 方法的代码参见文献[19],剩余方法使用了作者提供的代码。

表 1 是本文提出的方法与其他基线方法在 MIRFLICKR 和 NUSWIDE 这 2 个数据集上的 mAP@100 性能结果。表中 b 为比特位(bit),实验中最佳性能使用加粗字体表示。由表 1 可知本文所提出的方法性能达到较为显著的提升,但是 NUSWIDE 数据集上哈希码的精度变化并不明显,反而下降,这可能是由于设置的训练样本数量过少导致选取的类别样本分布不平衡。但是本文方法在全部的比特位上都取得了最佳性能,因此证实了该方法的有效性(HCOH 方法本身不能处理多标签数据集,为此去除了每个 chunk 的多标签数据,但是测试集和检索集是完整的,包含多标签数据)。

表 1 不同数据集上 mAP@100 的性能对比

Table 1 Performance comparison of mAP@100 on different datasets

对比 方法	MIRFLICKR					NUSWIDE				
	8 b	16 b	32 b	64 b	128 b	8 b	16 b	32 b	64 b	128 b
BSODH	0.586 3	0.586 5	0.597 9	0.597 7	0.597 9	0.567 0	0.569 6	0.566 7	0.568 1	0.566 5
HCOH	0.338 1	0.375 0	0.408 7	0.424 2	0.439 6	0.515 1	0.544 7	0.569 2	0.575 9	0.584 4
HMOH	0.575 5	0.574 8	0.585 6	0.592 8	0.599 4	0.491 3	0.512 0	0.529 6	0.536 1	0.548 4
FOH	0.564 8	0.551 2	0.527 5	0.586 9	0.578 1	0.555 2	0.595 4	0.571 9	0.576 1	0.554 2
OSEH	0.668 5	0.666 7	0.668 0	0.666 1	0.665 2	0.656 2	0.661 8	0.642 9	0.646 6	0.646 6

图 2 是本文提出的方法与基线方法在这 2 个数据集上的 Precision@100 的性能结果对比,由于使用细粒度标签嵌入与特征标签联合学习的策略促进语义信息的挖掘,从图中可以清晰地看到,本文所提出的方法超过了所有的对比方法。另外,在 NUSWIDE 数据集中可以观察到所有的方法都存在不同程

度的波动,产生此现象的原因可能是此数据集规模较大,涵盖的场景较多,内容复杂,因此产生了波动情况。由图2可知,本文方法在取得优越性能的同时,稳定性也较好,体现了本文方法在大规模数据集上的有效性。图3展示了mAP@all在32 b时随数据块的变化情况。从图3中可以看出,OSEH在MIR-FLICKR数据集中虽不是最优,但不同数据块之间性能相对平稳,同时取得了较好性能,证明了本文方法对新数据的适应性较好。在NUSWIDE数据集中不同数据块上,本文所提方法的mAP@all性能基本优于其他对比方法,此结果验证了本文所提方法适合用于在线检索的场景。

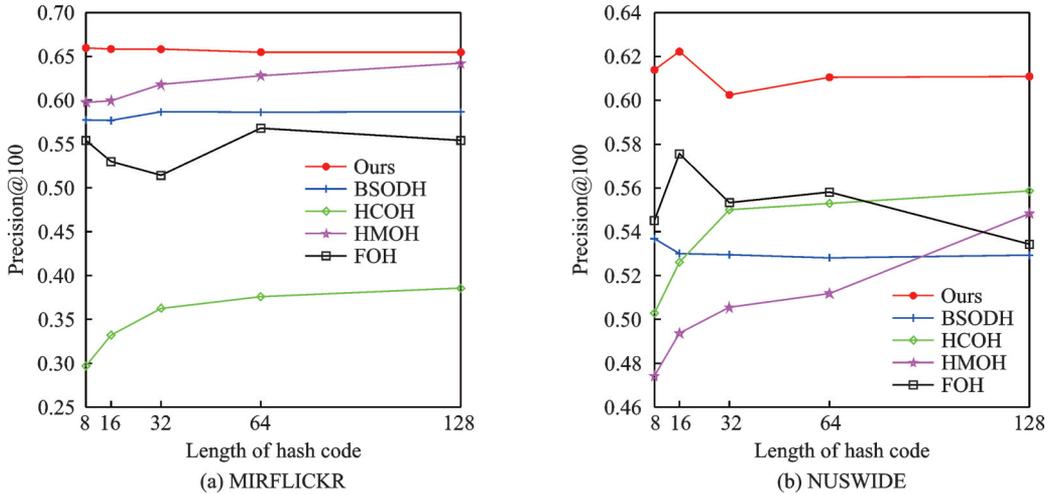


图2 各方法在两个数据集上 Precision@100 分数

Fig.2 Precision@100 of two datasets for different mthods

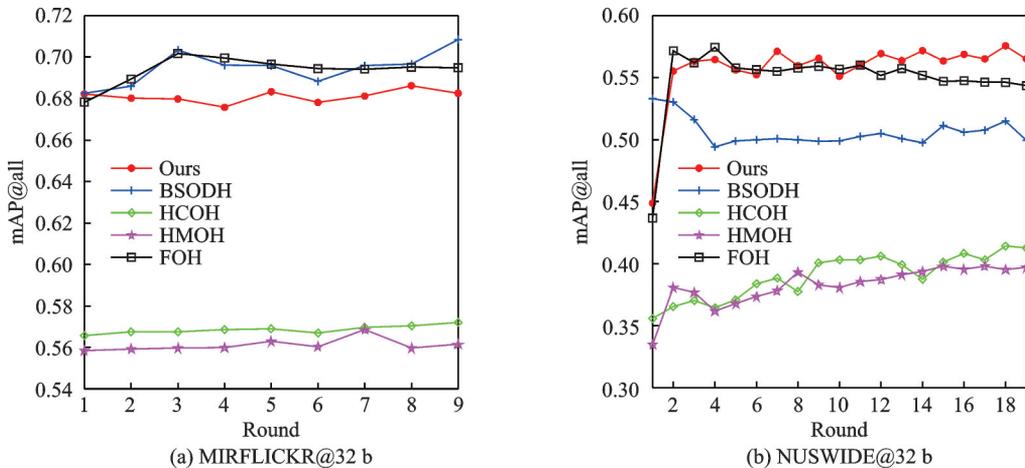


图3 各方法在两个数据集上 mAP@all 随数据块的变化曲线

Fig.3 mAP@all results varying with chunks of two datasets for different mthods

此外可通过人工主观评估进一步证明 OSEH 的有效性,具体可在数据集上随机选取 2 个索引,对比 OSEH 和其他方法 Top-10 的检索结果,由此能够在在线场景下检索得到与查询数据语义相关的样本。尽管可视化分析能够提供对模型表现的直观理解,但由于实验资源的限制,与所有基准模型的对比以及图文关系的详细展示过程需要较高的时间成本和资源消耗,综合考虑了实验的可行性和效能,本节未设置此部分实验。

3.3.2 消融实验

根据模型的设计方案,本文从3个角度验证所提方法的有效性。表2展示了消融实验的结果。

表2 不同数据集上模型消融实验(mAP@100)结果

Table 2 Performance of ablation study (mAP@100) on different datasets

对比方法	MIRFLICKR					NUSWIDE				
	8 b	16 b	32 b	64 b	128 b	8 b	16 b	32 b	64 b	128 b
OSEH	0.668 5	0.666 7	0.668 0	0.666 1	0.668 5	0.656 2	0.661 8	0.642 9	0.646 6	0.646 6
OSEH-x	0.656 1	0.648 3	0.654 8	0.656 0	0.657 0	0.516 4	0.598 4	0.605 3	0.582 6	0.611 8
OSEH-s	0.642 7	0.556 4	0.656 0	0.621 4	0.615 9	0.513 1	0.524 7	0.467 9	0.495 2	0.551 8
OSEH-b	0.659 2	0.661 0	0.657 0	0.657 3	0.657 5	0.632 1	0.628 5	0.628 1	0.616 2	0.616 2

OSEH-x是为了验证三元矩阵分解对特征学习的有效性。在此变体中去掉特征分解模块,可以看到性能有所下降,证明了将矩阵分解为正交分类映射、属性矩阵和标签矩阵,并将语义属性与特征标签联合学习来促进语义信息挖掘,对哈希码的学习有促进作用。

OSEH-s是为了验证相似性保持对哈希码的指导作用,此变体中去掉相似性保持模块,可以看到性能明显下降,证明了相似性保持对哈希码学习具有指导作用。

OSEH-b是为了在此变体中验证用 AL 指导哈希码是否具有有效性,由表2可见,在去掉该模块后,性能有所下降,因此可知属性矩阵与标签矩阵的结合可以作为细粒度标签更好地指导哈希码的学习。

消融实验的结果证明了特征和标签之间的互补学习可以实现语义增强,可以有效融合特征空间中的样本相似关系与独热码标签的类别区分信息,生成更具区分度的细粒度标签。

3.3.3 参数实验

此小节进行参数灵敏度分析,图4和图5分别展示了在两个数据集上参数 θ 和 σ 在 $10^{-5}\sim 10$ 之间取值时所取得的mAP@100的性能分数以及 k 和 α 在50~110与0.1~0.7之间取值时所取得的mAP@100的性能分数。 α 和 θ 在OSEH中是用来平衡特征矩阵分解与细粒度标签嵌入两个损失项, σ 是映射项的参数, k 用来选取合适的属性维度。从图4、5中可以看出:在MIRFLICKR数据集上本文方法随着参数改变性能变化相对来说比较稳定,证明在此数据集上本文方法相对鲁棒;在NUSWIDE数据集上当参

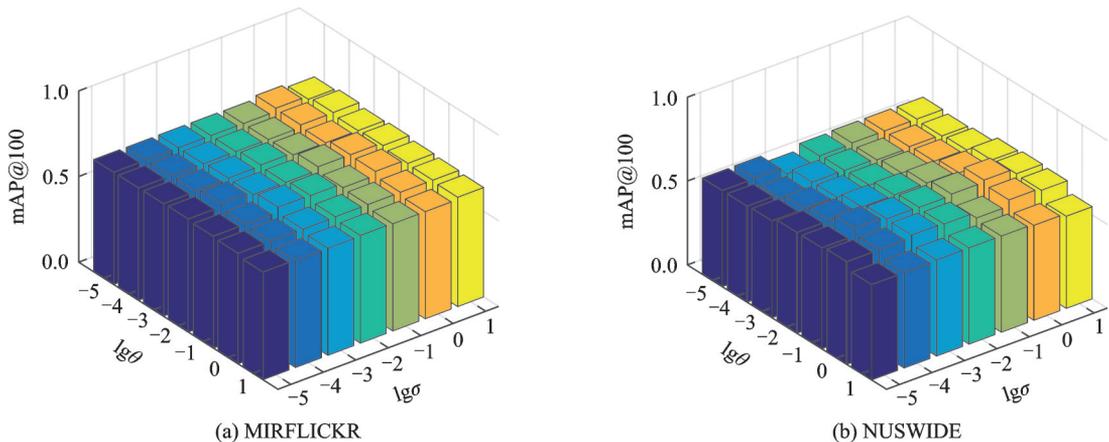
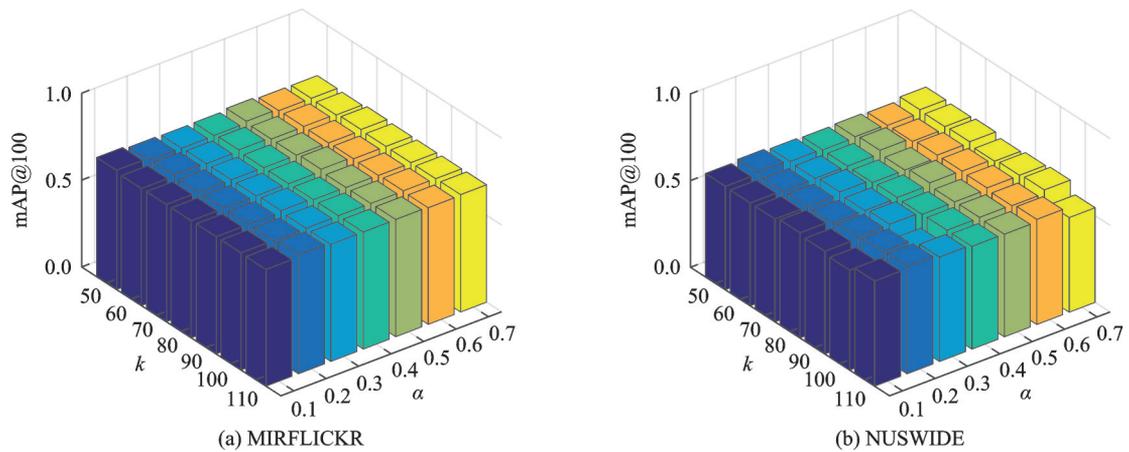


图4 两个数据集上参数 θ 和 σ 分析结果

Fig.4 Parameter sensitivity result for θ and σ on two datasets

图5 两个数据集上参数 k 和 α 分析结果Fig.5 Parameter sensitivity results for k and α on two datasets

数 θ 和 σ 在 $0.1\sim 1$ 之间取值时,本文方法会取得较好的性能, α 在 $0.5\sim 0.6$ 之间取值时, k 为100时取得较好性能。特征矩阵分解项的参数与属性维度结合,证明了属性辅助细粒度标签生成是有效的。

4 结束语

本文提出了一种面向语义增强的监督在线哈希学习方法。该方法通过强化监督信息的导向作用,有效促进了哈希码与哈希函数的学习。具体地,本文融合了特征空间中的样本相似性与标签空间中的类间区分性信息,生成细粒度的标签矩阵,从而解决了类间关系挖掘不足导致的流数据语义偏移问题。此外,将新生成的标签与成对相似关系嵌入到哈希学习中,有效缓解了累积数据的遗忘问题。在两个大规模数据集上的实验结果充分验证了所提方法的有效性。未来研究将致力于探索哈希码长的自适应策略,以更好地应对流数据中的类增长挑战。

参考文献:

- [1] CAO Yuan, QI Heng, ZHOU Wenrui, et al. Binary hashing for approximate nearest neighbor search on big data: A survey[J]. IEEE Access, 2017, 6: 2039-2054.
- [2] LIU Xingbo, LI Jiamin, NIE Xiushan, et al. Fast unsupervised cross-modal hashing with robust factorization and dual projection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, 20(12): 1-21.
- [3] CAKIR F, HE K, ADEL BARGAL S, et al. Mihash: Online hashing with mutual information[C]//Proceedings of the IEEE International Conference on Computer Vision. NJ: IEEE, 2017: 437-445.
- [4] ZHU Lei, ZHENG Chaoqin, LU Xu, et al. Efficient multi-modal hashing with online query adaption for multimedia retrieval[J]. ACM Transactions on Information Systems (TOIS), 2021, 40(2): 1-36.
- [5] 康潇,刘兴波,卢鹏宇,等.双重结构保持的在线跨模态哈希[J].计算机研究与发展, 2024, 61(11): 2923-2936. KANG Xiao, LIU Xingbo, LU Pengyu, et al. Online cross modal hashing with dual structure preservation[J]. Journal of Computer Research and Development, 2024, 61(11): 2923-2936.
- [6] LI Xuan, WU Wei, YUAN Yunhao, et al. Online unsupervised cross-view discrete hashing for large-scale retrieval[J]. Applied Intelligence, 2022, 52(13): 14905-14917.
- [7] 刘兴波,聂秀山,尹义龙.基于双向线性回归的监督离散跨模态散列方法[J].计算机研究与发展, 2020, 57(8): 1707-1714. LIU Xingbo, NIE Xiushan, YIN Yilong. Supervised discrete cross-modal hashing method based on two-way linear regression[J]. Journal of Computer Research and Development, 2020, 57(8): 1707-1714.
- [8] SU Ruoqi, WANG Di, ZHEN Huang, et al. Online adaptive supervised hashing for large-scale cross-modal retrieval[J]. IEEE Access, 2020, 8: 206360-206370.

- [9] FANG Yuzhi, LIU Li. Label projection online hashing for balanced similarity[J]. *Journal of Visual Communication and Image Representation*, 2021, 80: 103314.
- [10] FANG Yuzhi, LIU Li. Scalable supervised online hashing for image retrieval[J]. *Journal of Computational Design and Engineering*, 2021, 8(5): 1391-1406.
- [11] ZHENG Chaoqun, ZHU Lei, ZHANG Zheng, et al. LCEMH: Label correlation enhanced multi-modal hashing for efficient multi-modal retrieval[J]. *Information Sciences*, 2024, 659: 120064.
- [12] LIN Mingbao, JI Rongrong, CHEN Shen, et al. Similarity-preserving linkage hashing for online image retrieval[J]. *IEEE Transactions on Image Processing*, 2020, 29: 5289-5300.
- [13] YU Jun, WU Xiaojun, ZHANG Donglin. Unsupervised multi-modal hashing for cross-modal retrieval[J]. *Cognitive Computation*, 2022, 14(3): 1159-1171.
- [14] HUANG Longkai, YANG Qiang, ZHENG Weishi. Online hashing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(6): 2309-2322.
- [15] WENG Zhenyu, ZHU Yuesheng. Unsupervised online hashing with multi-bit quantization[C]//*Proceedings of the 16th ACCV Conference on Computer Vision*. Macao, China: [s.n.], 2022, 13847: 663-678.
- [16] CAKIR F, SCLAROFF S. Online supervised hashing[C]//*Proceedings of the IEEE International Conference on Image Processing*. Piscataway, NJ: IEEE, 2015: 2606-2610.
- [17] LIN Mingbao, JI Rongrong, LIU Hong, et al. Hadamard matrix guided online hashing[J]. *International Journal of Computer Vision*, 2020, 128(8): 2279-2306.
- [18] LIN Mingbao, JI Rongrong, LIU Hong, et al. Towards optimal discrete online hashing with balanced similarity[C]//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI, 2019: 8722-8729.
- [19] JIA Wenzhe, CAO Yuan, LIU Junwei, et al. Fast online hashing with multi-label projection[C]//*Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington, D C, USA: AAAI Press, 2023, 37(1): 1007-1014.
- [20] LI Li, SHU Zhenqiu, YU Zhengtao, et al. Robust online hashing with label semantic enhancement for cross-modal retrieval[J]. *Pattern Recognition*, 2024, 145:109972.
- [21] ZHANG Chongyu, LUO Xin, ZHAN Yuwei, et al. Self-distillation dual-memory online hashing with hash centers for streaming data retrieval[C]//*Proceedings of the 31st ACM International Conference on Multimedia*. [S.l.]: ACM, 2023: 6340-6349.
- [22] HUISKES M, LEW M. The MIRflickr retrieval evaluation[C]//*Proceedings of the 1st ACM SIGMM International Conference on Multimedia Information Retrieval*. New York: ACM, 2008: 9-43.
- [23] CHUA T S, TANG J, HONG R, et al. NUS-WIDE: A real-world web image database from national university of Singapore [C]//*Proceedings of the ACM International Conference on Image and Video Retrieval*. New York, USA: ACM, 2009: 1-9.
- [24] LIN Mingbao, JI Rongrong, LIU Hong, et al. Supervised online hashing via hadamard codebook learning[C]//*Proceedings of the 26th ACM International Conference on Multimedia*. [S.l.]: ACM, 2018: 1635-1643.

作者简介:



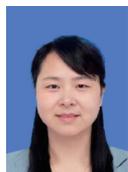
赵志杰(1999-),女,硕士研究生,研究方向:多媒体检索和计算机视觉,E-mail: zhaozhijie_01@163.com。



康潇(1997-),通信作者,女,博士研究生,研究方向:多媒体检索和计算机视觉,E-mail:sckx@mail.sdu.edu.cn。



张雪凝(1999-),女,硕士研究生,研究方向:哈希学习和弱监督学习。



王少华(1979-),女,博士,研究方向:多媒体检索和计算机视觉。



刘兴波(1995-),通信作者,男,博士,教授,研究方向:多媒体检索和计算机视觉,E-mail:sclxb@mail.sdu.edu.cn。



聂秀山(1981-),男,博士,教授,研究方向:多媒体检索、多媒体安全和计算机视觉。

(编辑:刘彦东)