

半配对的多模态询问哈希方法

庾骏^① 马江涛^① 咸 阳^① 侯瑞霞*^② 孙 伟^③

^①(郑州轻工业大学计算机与通信工程学院 郑州 450000)

^②(中国林业科学研究院资源信息研究所 北京 100091)

^③(中国农业科学院农业信息研究所 北京 100081)

摘要: 多模态哈希能够将异构的多模态数据转化为联合的二进制编码串。由于其具有低存储成本、快速的汉明距离排序的优点,已经在大规模多媒体检索中受到了广泛的关注。现有的多模态哈希方法假设所有的询问数据都具备完整的多种模态信息以生成它们的联合哈希码。然而,实际应用中很难获得全完整的多模态信息,针对存在模态信息缺失的半配对询问场景,该文提出一种新颖的半配对询问哈希(SPQH),以解决半配对的询问样本的联合编码问题。首先,提出的方法执行投影学习和跨模态重建学习以保持多模态数据间的语义一致性。然后,标签空间的语义相似结构信息和多模态数据间的互补信息被有效地捕捉以学习判别性的哈希函数。在询问编码阶段,通过学习到的跨模态重构矩阵为未配对的样本数据补充缺失的模态特征,然后再经习得的联合哈希函数生成哈希特征。相比最先进的基线方法,在Pascal Sentence, NUS-WIDE和IAPR TC-12数据集上的平均检索精度提高了2.48%。实验结果表明该算法能够有效编码半配对的多模态询问数据,取得了优越的检索性能。

关键词: 多模态信息检索; 哈希; 半配对数据; 跨模态重建; 二值化编码

中图分类号: TN911.7; TP391

文献标识码: A

文章编号: 1009-5896(2024)02-0481-11

DOI: [10.11999/JEIT231072](https://doi.org/10.11999/JEIT231072)

Semi-paired Multi-modal Query Hashing Method

YU Jun^① MA Jiangtao^① XIAN Yang^① HOU Ruixia^② SUN Wei^③

^①(College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China)

^②(Research Institute of Resource Information Techniques, CAF, Beijing, 100091, China)

^③(Agricultural Information Institute of CAAS, Beijing 100081, China)

Abstract: Multimodal hashing can convert heterogeneous multimodal data into unified binary codes. Due to its advantages of low storage cost and fast Hamming distance sorting, it has attracted widespread attention in large-scale multimedia retrieval. Existing multimodal hashing methods assume that all query data possess complete multimodal information to generate their joint hash codes. However, in practical applications, it is difficult to obtain fully complete multimodal information. To address the problem of missing modal information in semi-paired query scenarios, a novel Semi-paired Query Hashing (SPQH) method is proposed to solve the joint encoding problem of semi-paired query samples. Firstly, the proposed method performs projection learning and cross-modal reconstruction learning to maintain semantic consistency among multimodal data. Then, the semantic similarity structure information of the label space and complementary information among multimodal data are effectively captured to learn a discriminative hash function. During the query encoding stage, the missing modal features of unpaired sample data are completed using the learned cross-modal reconstruction matrix, and then the hash features are generated using the learned joint hash function. Compared to state-of-the-art baseline methods, the average retrieval accuracy on the Pascal Sentence, NUS-WIDE, and IAPR TC-12 datasets has improved by 2.48%. Experimental results demonstrate that the algorithm can effectively encode

收稿日期: 2023-10-08; 改回日期: 2024-01-31; 网络出版: 2024-01-31

*通信作者: 侯瑞霞 houreix@ifrit.ac.cn

基金项目: 国家自然科学基金(32271880), 河南省科技攻关项目基金(222102210064), 河南省自然科学基金(232300420150)

Foundation Items: The National Natural Science Foundation of China (32271880), The Science and Technology Research Project of Henan Provincial Department (222102210064), The Natural Science Foundation of Henan Province Science (232300420150)

semi-paired multimodal query data and achieve superior retrieval performance.

Key words: Multimodal retrieval; Hashing; Semi-paired data; Cross-modal reconstruction; Binary codes

1 引言

目前, 社交网络中的多媒体数据持续以惊人的速度快速增长^[1], 使得信息搜索遇到前所未有的巨大挑战。哈希作为一种有效的特征表示技术, 以其低存储、高效率的优点在信息检索、计算机视觉等领域发挥着重要作用^[2]。哈希方法将高维数据映射成一串紧凑的二进制代码, 然后通过快速异或运算来加速大规模数据中的最近邻(Approximate Nearest Neighbor, ANN)搜索。先驱者们专注于单一模态搜索而提出了众多单模态哈希搜索方法^[3-11], 例如迭代量化(ITerative Quantization, ITQ)^[3]通过学习零中心化后的数据的正交旋转矩阵, 降低了从数据空间到二元超立方体空间过程的量化误差。离散局部线性嵌入哈希(Discrete Locality Linear embedding Hashing, DLLH)^[6]将原始数据的局部流形结构保持到哈希码中, 并且提供了一种基于锚点的加速方案, 将其扩展到大规模数据搜索。基于哈达玛密码本的在线哈希(Hadamard codebook based online hashing, HCOH)^[8]将共享相同标签的数据与哈达玛矩阵中的码字相关联, 以在线方式训练哈希函数。单模哈希广泛应用于图像搜索并取得令人满意的性能。

然而, 在实际搜索任务中, 目标对象通常由来源不同的多种模态数据所描述, 例如微信朋友圈发布的图片往往会附有相关的文字描述内容。一种直观的方式是扩展单模哈希, 将多模态数据简单拼接为统一的高维特征来处理。然而, 这种处理方式可能会造成了信息冗余和维数灾难, 因此单模态哈希方法在多模态任务中表现不佳。为了克服这一问题, 一些研究者们致力于研究多种模态特征的联合哈希编码问题并提出了众多多模态哈希方法^[12-19], 它们将异构多模态数据转换成一串联合的二进制编码。例如多特征哈希(Multiple Feature Hashing, MFH)^[12], 多视图潜在哈希(Multi-view Latent Hashing, MVLH)^[13], 多视图对齐哈希(Multiview Alignment Hashing, MAH)^[14], 深度多模哈希(Deep Multi-modal Hashing, SIDMH)^[15], 哈达玛矩阵引导的多模哈希(Hadamard matrix-Guided Multi-modal Hashing, HGMH)^[16], 基于松弛Hadamard矩阵的多模态融合哈希^[17], 具有动态查询自适应的在线多模态哈希(Online Multi-modal Hashing with Dynamic Query-adaption, OMH-DQ)^[18]和自适应多模态融合哈希(Adaptive Multi-

modal Fusion Hashing, AMFH)^[19]等。多模态哈希方法旨在为完全配对的多模态样本生成联合的哈希编码, 获得的联合哈希码融合了多个模态之间丰富的互补信息, 相比单模态哈希方法而言, 具有更好的表征能力。然而, 由于习得的联合哈希函数考虑了所有的模态信息, 因此这些方法对新的询问样本编码时需要确保样本中的模态信息完整。事实上, 在询问阶段并不能保证所有的新样本数据都为完全配对的多模态数据, 部分样本存在缺失某些模态信息的可能, 本文将这种场景称之为半配对的多模态询问场景。针对这个上述问题, 一些半配对的跨模态哈希方法, 例如半配对哈希(Semi-Paired Hashing, SPH)^[20], 半配对半监督多模态哈希(Semi-paired and Semi-supervised Multi-modal Hashing, SSMH)^[21], 未配对的跨模态哈希(Unpaired Cross-Modal Hashing, UCMH)^[22]和基于自训练的半监督和半配对哈希(Self- Training based Cross-modal Hashing, STCH)^[23]被相继提出, 它们为每个模态数据学习一个独立的哈希函数。但是, 这些方法的多个独立哈希函数存在差异性, 很难实现多模态的联合哈希编码。本文提出了一种半配对多模态询问哈希(Semi-Paired Query Hashing, SPQH), 为半配对的多模态询问生成联合的哈希编码提供了有效的解决方案。图1展示了本文提出的SPQH的流程图。所提框架包括训练阶段和询问编码阶段。在训练阶段, 联合执行投影学习、跨模态重构学习、多模态融合学习和语义保持哈希学习来学习联合的判别性哈希函数。在询问编码阶段, 对于模态缺失的询问样本, 提出的模型巧妙地利用学习到的交叉重构矩阵来补全缺失的模态特征, 然后再融合多模态特征以获得询问实例的联合哈希码。本文提出的SPQH的主要贡献总结如下:

(1) 开发了一种多模态联合哈希学习框架, 将子空间学习和哈希学习结合成一个统一的学习架构。提出的模型融合了多模态信息并保留标签空间中的语义结构信息以学习具有判别性的联合哈希函数。

(2) 设计的跨模态重构项增强了异质模态之间的语义一致性表示, 能够利用学习到的重构矩阵有效地解决了询问数据中模态不完整的样本的联合哈希编码问题。

(3) 提出了一种有效的优化方法来求解带有离散约束的目标函数。在3个公开数据集上的评估结果表明了本文所提出的方法优于最新的多模态哈希方法。

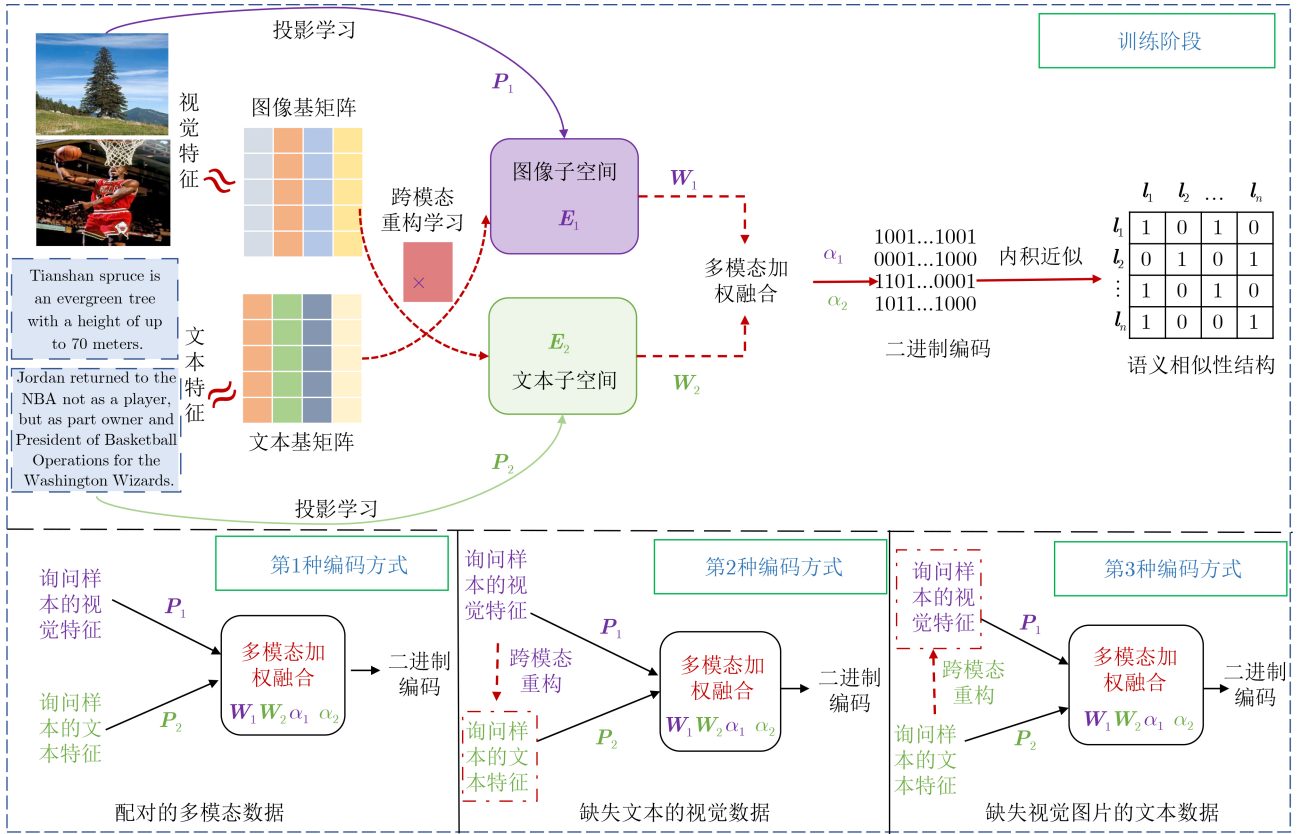


图 1 SPQH框架的示意图

在行文结构上, 本文剩余部分安排如下。第2节详细描述提出的半配对的多模态询问哈希模型和优化方法。第3节展示并讨论了在3个数据集上的实验结果。第4部分得出论文的结论。

2 半配对的多模态询问哈希方法

2.1 模型阐述

本文以视觉图片和语言文本两个模态为例进行多模态联合哈希学习建模, 假设给定数据集由 n 个图像-文本样本组成。图像集和文本集分别表示为 $\mathbf{Y}^{(1)} = [\mathbf{y}_1^{(1)}, \mathbf{y}_2^{(1)}, \dots, \mathbf{y}_n^{(1)}]$ 和 $\mathbf{Y}^{(2)} = [\mathbf{y}_1^{(2)}, \mathbf{y}_2^{(2)}, \dots, \mathbf{y}_n^{(2)}]$, 其中 $\mathbf{y}_i^{(1)} \in R^{d_x}$ 和 $\mathbf{y}_i^{(2)} \in R^{d_y}$, 通常 $d_x \neq d_y$ 。标签矩阵 $\mathbf{L} = [l_1, l_2, \dots, l_n] \in R^{r \times n}$, 其中 r 表示类别数。 $l_i = [l_{i1}, l_{i2}, \dots, l_{ir}]^T \in R^r$, 当第 i 个样本被划分为第 j 类时则 $l_{ij} = 1$, 否则 $l_{ij} = 0$ 。本文采用径向基核函数获取图像特征和文本特征, 即第 m 个模态中的第

$$i \text{ 个样本的核特征为 } \mathbf{X}_m^i = \left[\exp \left(\frac{\|\mathbf{y}_i^{(m)} - \mathbf{A}_1^{(m)}\|^2}{-2\sigma_m^2} \right), \exp \left(\frac{\|\mathbf{y}_i^{(m)} - \mathbf{A}_2^{(m)}\|^2}{-2\sigma_m^2} \right), \dots, \exp \left(\frac{\|\mathbf{y}_i^{(m)} - \mathbf{A}_k^{(m)}\|^2}{-2\sigma_m^2} \right) \right]^T,$$

其中 σ_m 是第 m 个模态对应的核宽, ($m=1, 2$);

$\mathbf{A}^{(1)} \in R^{d_x \times k}$ 和 $\mathbf{A}^{(2)} \in R^{d_y \times k}$ 分别是从小图模态和文本模态中随机选择的锚点样本集, 其中 k 是锚点样本的数目。本文使用 \mathbf{X}_1 和 \mathbf{X}_2 分别表示经过核化计算后的图像模态特征矩阵和文本模态特征矩阵。

样本中的不同模态描述相同的语义内容, 成对的不同模态之间的语义一致性可以作为跨越不同模态的桥梁。某个模态可以通过其他模态的子空间重建, 以捕获模态之间的相关性。因此, 通过找到一个潜在的子空间, 其中任意模态数据通过其他模态的子空间特征来重构。上述想法转换为式(1)。

$$\min_{\mathbf{P}_1, \mathbf{P}_2, \mathbf{E}_1, \mathbf{E}_2, \mathbf{U}_1, \mathbf{U}_2} \lambda \left(\|\mathbf{P}_1 \mathbf{X}_1 - \mathbf{E}_1\|_F^2 + \|\mathbf{X}_1 - \mathbf{U}_1 \mathbf{E}_2\|_F^2 \right) + (1 - \lambda) \cdot \left(\|\mathbf{P}_2 \mathbf{X}_2 - \mathbf{E}_2\|_F^2 + \|\mathbf{X}_2 - \mathbf{U}_2 \mathbf{E}_1\|_F^2 \right) \quad (1)$$

其中 $\mathbf{P}_m \in R^{d \times k}$ 和 $\mathbf{U}_m \in R^{k \times d}$ 分别表示第 m 个模态的映射矩阵和基矩阵。 $\mathbf{E}_m \in R^{d \times n}$ 表示第 m 个模态在子空间中的特征表示。异构模态之间的语义内容是一致的, 且信息上存在互补性, 因此, 本文通过融合互补的多模态信息并且保留标签空间中的语义结构, 有助于增强哈希模型的判别性。具体而言, 问题表述为式(2)。

$$\begin{aligned} & \min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{B}, \alpha_m} \sum_{m=1}^2 (\alpha_m)^t \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2 \\ & + \beta \|\mathbf{S} - \mathbf{B}^T \mathbf{B}\|_F^2 \\ \text{s.t. } & \mathbf{B} \in \{+1, -1\}^{c \times n}, \sum_{m=1}^2 \alpha_m = 1, \alpha_m > 0 \end{aligned} \quad (2)$$

其中 $\mathbf{W}_m \in R^{c \times d}$ 表示第 m 个模式的映射矩阵, α_m 表示第 m 个模式的加权系数, t 是平滑参数. $\mathbf{S} \in R^{n \times n}$ 表示语义近邻矩阵, 被定义为 $S_{ij} = \frac{l_i l_j}{\|l_i\|_2 \|l_j\|_2}$. 联合式(1)和式(2)后, 整体的优化问题被定义成式(3).

$$\begin{aligned} & \min_{\mathbf{P}_1, \mathbf{P}_2, \mathbf{E}_1, \mathbf{E}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{W}_1, \mathbf{W}_2, \mathbf{B}, \alpha_m} \\ & \lambda (\|\mathbf{P}_1 \mathbf{X}_1 - \mathbf{E}_1\|_F^2 + \|\mathbf{X}_1 - \mathbf{U}_1 \mathbf{E}_2\|_F^2) \\ & + (1 - \lambda) (\|\mathbf{P}_2 \mathbf{X}_2 - \mathbf{E}_2\|_F^2 + \|\mathbf{X}_2 - \mathbf{U}_2 \mathbf{E}_1\|_F^2) \\ & + \sum_{m=1}^2 (\alpha_m)^t \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2 + \beta \|\mathbf{S} - \mathbf{B}^T \mathbf{B}\|_F^2 \\ & + \gamma \Phi(\mathbf{P}_1, \mathbf{P}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{W}_1, \mathbf{W}_2) \\ \text{s.t. } & \mathbf{B} \in \{+1, -1\}^{c \times n}, \sum_{m=1}^2 \alpha_m = 1, \alpha_m > 0 \end{aligned} \quad (3)$$

其中 $\Phi(\mathbf{P}_1, \mathbf{P}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{W}_1, \mathbf{W}_2) = \|\mathbf{P}_1\|_F^2 + \|\mathbf{P}_2\|_F^2 + \|\mathbf{U}_1\|_F^2 + \|\mathbf{U}_2\|_F^2 + \|\mathbf{W}_1\|_F^2 + \|\mathbf{W}_2\|_F^2$ 是避免过拟合的正则化项, λ, β 和 γ 是平衡参数.

2.2 模型的优化方法

由于式(3)是一个多变量的非凸优化问题, 直接求解比较困难. 在本小节中, 采用交替优化的方式来优化这些变量, 通过固定其他变量, 每次优化一个变量, 更新规则总结为以下9个步骤.

步骤1 其他变量不变, 关于 \mathbf{P}_1 的子目标函数为

$$\min_{\mathbf{P}_1} \lambda \|\mathbf{P}_1 \mathbf{X}_1 - \mathbf{E}_1\|_F^2 + \gamma \|\mathbf{P}_1\|_F^2 \quad (4)$$

令式(4)关于 \mathbf{P}_1 的导数为零, 得到了 \mathbf{P}_1 的闭合解

$$\mathbf{P}_1 = \lambda \mathbf{E}_1 \mathbf{X}_1^T (\lambda \mathbf{X}_1 \mathbf{X}_1^T + \gamma \mathbf{I})^{-1} \quad (5)$$

步骤2 其他变量不变, 关于 \mathbf{P}_2 的目标函数为

$$\min_{\mathbf{P}_2} \lambda \|\mathbf{P}_2 \mathbf{X}_2 - \mathbf{E}_2\|_F^2 + \gamma \|\mathbf{P}_2\|_F^2 \quad (6)$$

然后, 令式(6)关于 \mathbf{P}_2 的导数为零, 得到

$$\mathbf{P}_2 = \lambda \mathbf{E}_2 \mathbf{X}_2^T (\lambda \mathbf{X}_2 \mathbf{X}_2^T + \gamma \mathbf{I})^{-1} \quad (7)$$

步骤3 其他变量不变, 关于 \mathbf{E}_1 的目标函数为

$$\begin{aligned} & \min_{\mathbf{E}_1} \lambda \|\mathbf{P}_1 \mathbf{X}_1 - \mathbf{E}_1\|_F^2 + (1 - \lambda) \|\mathbf{X}_2 - \mathbf{U}_2 \mathbf{E}_1\|_F^2 \\ & + (\alpha_1)^t \|\mathbf{W}_1 \mathbf{E}_1 - \mathbf{B}\|_F^2 \end{aligned} \quad (8)$$

令式(8)关于 \mathbf{E}_1 的导数为零, 可以得到

$$\begin{aligned} \mathbf{E}_1 = & (\lambda \mathbf{I} + (1 - \lambda) \mathbf{U}_2^T \mathbf{U}_2 + \alpha_1^t \mathbf{W}_1^T \mathbf{W}_1)^{-1} \\ & (\lambda \mathbf{P}_1 \mathbf{X}_1 + (1 - \lambda) \mathbf{U}_2^T \mathbf{X}_2 + \alpha_1^t \mathbf{W}_1^T \mathbf{B}) \end{aligned} \quad (9)$$

步骤4 其他变量不变, 关于 \mathbf{E}_2 的目标函数为

$$\begin{aligned} & \min_{\mathbf{E}_2} (1 - \lambda) \|\mathbf{P}_2 \mathbf{X}_2 - \mathbf{E}_2\|_F^2 + \lambda \|\mathbf{X}_1 - \mathbf{U}_1 \mathbf{E}_2\|_F^2 \\ & + (\alpha_2)^t \|\mathbf{W}_2 \mathbf{E}_2 - \mathbf{B}\|_F^2 \end{aligned} \quad (10)$$

令式(10)关于 \mathbf{E}_2 的导数为零, 可以得到

$$\begin{aligned} \mathbf{E}_2 = & ((1 - \lambda) \mathbf{I} + \lambda \mathbf{U}_1^T \mathbf{U}_1 + \alpha_2^t \mathbf{W}_2^T \mathbf{W}_2)^{-1} \\ & \cdot ((1 - \lambda) \mathbf{P}_2 \mathbf{X}_2 + \lambda \mathbf{U}_1^T \mathbf{X}_1 + \alpha_2^t \mathbf{W}_2^T \mathbf{B}) \end{aligned} \quad (11)$$

步骤5 其他变量不变, 关于 \mathbf{U}_1 的目标函数为

$$\min_{\mathbf{U}_1} \lambda \|\mathbf{X}_1 - \mathbf{U}_1 \mathbf{E}_2\|_F^2 + \gamma \|\mathbf{U}_1\|_F^2 \quad (12)$$

令式(12)关于 \mathbf{U}_1 的导数为零, 可以得到

$$\mathbf{U}_1 = \lambda \mathbf{X}_1 \mathbf{E}_2^T (\lambda \mathbf{E}_2 \mathbf{E}_2^T + \gamma \mathbf{I})^{-1} \quad (13)$$

步骤6 其他变量不变, 关于 \mathbf{U}_2 的目标函数为

$$\min_{\mathbf{U}_2} (1 - \lambda) \|\mathbf{X}_2 - \mathbf{U}_2 \mathbf{E}_1\|_F^2 + \gamma \|\mathbf{U}_2\|_F^2 \quad (14)$$

令式(14)关于 \mathbf{U}_2 的导数为零, 可以得到

$$\mathbf{U}_2 = (1 - \lambda) \mathbf{X}_2 \mathbf{E}_1^T ((1 - \lambda) \mathbf{E}_1 \mathbf{E}_1^T + \gamma \mathbf{I}) \quad (15)$$

步骤7 其他变量不变, 关于 \mathbf{W}_m 的目标函数为

$$\min_{\mathbf{W}_m} (\alpha_m)^t \sum_{m=1}^2 \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2 + \gamma \sum_{m=1}^2 \|\mathbf{W}_m\|_F^2 \quad (16)$$

令式(16)关于 \mathbf{W}_m 的导数为零, 可以得到

$$\mathbf{W}_1 = (\alpha_1)^t \mathbf{B} \mathbf{E}_1^T ((\alpha_1)^t \mathbf{E}_1 \mathbf{E}_1^T + \gamma \mathbf{I})^{-1} \quad (17)$$

$$\mathbf{W}_2 = (\alpha_2)^t \mathbf{B} \mathbf{E}_2^T ((\alpha_2)^t \mathbf{E}_2 \mathbf{E}_2^T + \gamma \mathbf{I})^{-1} \quad (18)$$

步骤8 其他变量不变, 关于 α_m 的目标函数为

$$\begin{aligned} & \min_{\alpha_m} \sum_{m=1}^2 G_m \\ \text{s.t. } & \sum_{m=1}^2 \alpha_m = 1, \alpha_m > 0 \end{aligned} \quad (19)$$

其中 $G_m = \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2$, 通过优化式(19)的拉格朗日函数得到最优 α_m

$$\alpha_m = \frac{(G_m)^{\frac{1}{1-t}}}{\sum_{m=1}^2 (G_m)^{\frac{1}{1-t}}} \quad (20)$$

步骤9 其他变量不变, 关于 \mathbf{B} 的目标函数为

$$\begin{aligned} & \min_{\alpha_m} \sum_{m=1}^2 \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2 + \beta \|\mathbf{B}^T \mathbf{B} - \mathbf{S}\|_F^2 \\ \text{s.t. } & \mathbf{B} \in \{+1, -1\}^{c \times n} \end{aligned} \quad (21)$$

带离散约束的式(21)是一个NP-hard问题。通过引入辅助变量 \mathbf{V} ，式(21)转化为它的等价形式

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{V}} \sum_{m=1}^2 \|\mathbf{W}_m \mathbf{E}_m - \mathbf{B}\|_F^2 + \beta \|\mathbf{B}^T \mathbf{V} - \mathbf{S}\|_F^2 \\ + \rho \|\mathbf{B} - \mathbf{V}\|_F^2 \\ \text{s.t. } \mathbf{B} \in \{+1, -1\}^{c \times n} \end{aligned} \quad (22)$$

其中 ρ 是一个平衡参数。采用交替优化方法求解 \mathbf{B} 和 \mathbf{V} 。具体地，通过固定 \mathbf{V} ，可以得到 \mathbf{B} 的闭合解

$$\mathbf{B} = \text{sign} \left(\sum_{m=1}^2 (\alpha_m)^t \mathbf{W}_m \mathbf{E}_m + \beta \mathbf{V} \mathbf{S}^T + \rho \mathbf{V} \right) \quad (23)$$

保持 \mathbf{B} 不变时，令目标函数(22)关于 \mathbf{V} 的导数为零，可以得到

$$\mathbf{V} = (\beta \mathbf{B} \mathbf{B}^T + \rho \mathbf{I})^{-1} (\beta \mathbf{B} \mathbf{S} + \rho \mathbf{B}) \quad (24)$$

在联合学习过程中，所有变量将被迭代性地交替优化，直到总体目标函数收敛或迭代次数达到预设的最大值，整个优化过程被总结在算法1中。通过算法1可获得式(3)中最优的 $\mathbf{P}_1, \mathbf{P}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{W}_1$ 和 \mathbf{W}_2 值。这些求得的最优变量将被传递给后续询问编码阶段，用于为多模态询问样本生成联合的哈希编码。

2.3 非成对的询问扩展和联合哈希编码阶段

现有的多模态哈希方法假设所有询问数据都具有完整的模态信息。然而，在半配对的多模态数据应用中存在一部分不完整模态的询问实例，即某些实例样本有模态缺失的情形。本小节利用2.2节中习得的 \mathbf{U}_1 和 \mathbf{U}_2 来处理未成对的询问数据。具体来说，令 \mathbf{X}_1^q 表示第 q 个询问样本的视觉图片模态， \mathbf{X}_2^q 表示第 q 个询问样本的语言文本模态，如果第 q 个询问样本的图像模态缺失，则通过已知的文本模态特征 \mathbf{X}_2^q 重构缺失的图像特征。如

算法 1 半配对的多模态询问哈希模型

输入：训练集 $\mathbf{O} = (\mathbf{y}_1^{(1)}, \mathbf{y}_1^{(2)}, l_1), (\mathbf{y}_2^{(1)}, \mathbf{y}_2^{(2)}, l_2), \dots, (\mathbf{y}_n^{(1)}, \mathbf{y}_n^{(2)}, l_n)$ ，核特征表示矩阵： $\mathbf{X}_1, \mathbf{X}_2$ 。

输出： $\mathbf{P}_1, \mathbf{P}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{W}_1, \mathbf{W}_2, \alpha_m$ 。

- ① 初始化 $\mathbf{U}_1, \mathbf{U}_2, \mathbf{E}_1, \mathbf{E}_2, \mathbf{W}_1, \mathbf{W}_2, \mathbf{B}, \alpha_1, \alpha_2$ 。
- ② for iter =1: ξ do
- ③ 根据等式(5)和式(7)分别更新 \mathbf{P}_1 和 \mathbf{P}_2 ；
- ④ 根据等式(9)和式(11)分别更新 \mathbf{E}_1 和 \mathbf{E}_2 ；
- ⑤ 根据等式(13)和式(15)分别更新 \mathbf{U}_1 和 \mathbf{U}_2 ；
- ⑥ 根据等式(17)和式(18)分别更新 \mathbf{W}_1 和 \mathbf{W}_2 ；
- ⑦ 根据式(20)更新 α_1 和 α_2 ；
- ⑧ 根据式(23)和式(24)更新 \mathbf{B} ；
- ⑨ end for

$$\min_{\mathbf{X}_1^q} \|\mathbf{X}_1^q - \mathbf{U}_1 \mathbf{E}_2^q\|_F^2 \quad (25)$$

其中 \mathbf{E}_2^q 为文本模态的子空间表示矩阵。可以很容易地获取图像模态特征

$$\mathbf{X}_1^q = \mathbf{U}_1 \mathbf{P}_2 \mathbf{X}_2^q \quad (26)$$

同样，可以依据已知的图像模态特征 \mathbf{X}_1^q 对缺失的文本模态特征进行补充，具体如

$$\min_{\mathbf{X}_2^q} \|\mathbf{X}_2^q - \mathbf{U}_2 \mathbf{E}_1^q\|_F^2 \quad (27)$$

其中 \mathbf{E}_1^q 为图像模态的子空间表示矩阵。计算的文本模态特征

$$\mathbf{X}_2^q = \mathbf{U}_2 \mathbf{P}_1 \mathbf{X}_1^q \quad (28)$$

多模态实例中的多模态信息具有很强的互补性，习得的哈希函数能有效地融合多个模态数据以获得联合的哈希表示特征，第 q 个询问样本的联合哈希特征表示成 $b_q = \text{sign}(\alpha_1 \mathbf{W}_1 \mathbf{P}_1 \mathbf{X}_1^q + \alpha_2 \mathbf{W}_2 \mathbf{P}_2 \mathbf{X}_2^q)$ 。

2.4 算法复杂性分析

算法1通过迭代更新每个变量来最小化目标函数。在本小节中，将阐述更新算法的计算复杂度。在训练阶段，更新 \mathbf{P}_1 和 \mathbf{P}_2 的时间复杂度为 $\mathcal{O}(dnk + k^3)$ ，更新 $\mathbf{E}_1, \mathbf{E}_2, \mathbf{U}_1$ 和 \mathbf{U}_2 的成本为 $\mathcal{O}(dnk)$ ，更新 \mathbf{W}_1 和 \mathbf{W}_2 需要消耗 $\mathcal{Q}(nd^2)$ 。当更新 \mathbf{B} 时，如果直接使用 \mathbf{S} 进行优化计算，时间复杂度为 $\mathcal{O}(n^2c)$ 。本文使用标签矩阵 $\mathbf{L}^T \mathbf{L}$ 代替 \mathbf{S} ，计算复杂度降为 $\mathcal{O}(cdn)$ 。假设算法在 ξ 次迭代后收敛，所提方法的总复杂度为 $\mathcal{O}(\xi(dkn + k^3))$ 。

3 实验

3.1 实验设置

为了验证提出的方法的性能，本节将在Pascal Sentence^[24]、NUS-WIDE^[25]和IAPR TC-12^[26]等3个基准数据集上与多种主流的哈希方法进行对比实验，即使用测试集数据从检索集中检索相关结果，文中采用平均查准率(mean Average Precision, mAP)指标衡量检索性能，越大的mAP值代表越好的检索性能。3个数据集的详细统计如表1所示，实

表 1 3个基准数据集的统计数据

数据集	Pascal Sentence	NUS-WIDE	IAPR TC-12
数据集的大小	1 000	186 577	20 000
训练集大小	600	5 000	5 000
检索集大小	600	186 577	18 000
测试集大小	400	1 866	2 000
类别数目	20	10	255

验中的询问数据指的是除训练集以外的数据, 包含了测试集和检索集。在对比实验中, 将提出的SPQH与单模态哈希方法(包含ITQ^[3], LSH^[4], DLLH^[6], HCOH^[8]), 多模态哈希方法(包含MFH^[12], MVLH^[13], OMH-DQ^[18], SIDMH^[15], AMFH^[19]), 半配对跨模态哈希模型(SPH^[20], SSMH^[21], UCMH^[22], STCH^[23])和深度学习方法(DMHOR^[27], DMVH^[28], FGCMH^[29])进行性能比较。在实验中, 完全配对的询问场景指的是在测试过程中提供了完全配对的询问集, 而半配对的询问场景说明询问集中有一定比例的未配对的询问样本, 其中完全未配对的询问场景表示询问集仅含有图片模态或者文本模态。由于SPQH方法在训练阶段是基于完全配对的图文样本来训练的, 所以为了进行公平比较, 单模态哈希方法(ITQ, LSH, DLLH和HCOH)将多个模态输入特征串联起来进行模型训练, 半配对跨模态哈希模型(SPH, SSH, SSMH, UCMH, STCH)将配对样本的比例设置成100%来完成模型训练。基线方法中的超参数根据其原始论文中给出的设置建议进行设置。提出的SPQH通过固定其他参数, 在较大范围内调整方法中的每个超参数以寻找其最优设置。在对比实验中, 详细参数设置如下: 在Pascal Sentence数据集上, $\lambda = 0.5$, $\beta = 1e^5$, $\gamma = 1e^{-3}$, $\rho = 1e^3$; 在NUS-WIDE数据集上 $\lambda = 0.1$, $\beta = 1e^1$, $\gamma = 1e^{-3}$, $\rho = 1e^3$; 在IAPR TC-12数据集上, $\lambda = 0.9$, $\beta = 1e^3$, $\gamma = 1e^{-5}$, $\rho = 1e^{-5}$ 。

3.2 检索精度比较

本节将呈现在3个数据集上所有方法在不同哈希码长度下的检索精度, 哈希编码长度分别被设置成16位、32位、64位和128位, 以全面地评估所提方法的性能。在测试阶段, 执行图像查询图像、文本查询文本、图像-文本实例查询图像-文本实例3种

检索过程, 分别简称为“I2I”, “T2T”和“O2O”。文中SPQH后接的后缀标识(如“all”, “img”, “txt”)是为了区分给定的询问数据中包含的模态类型。

(1) Pascal Sentence数据集上的对比实验: 表2展现了所有对比方法在Pascal Sentence数据集上的mAP值。从对比的实验结果中可以观察到SPQH_all优于所有的单模态哈希方法。当码长为16 bit时, SPQH_all的性能劣于多模态哈希方法SIDMH和AMFH, 而在码长大于32 bit时, 比次优的多模态哈希方法AMFH平均提高了2%。如图2(a)所示, 通过改变汉明半径以返回落在不同半径内的点, 绘制Pascal Sentence数据集上所有对比方法的PR曲线, 本文提出的SPQH性能优于所有基线方法。在表3中, 可以观察到本文提出的SPQH在Pascal Sentence数据集上一致地超越所有半配对的跨模态哈希方法, 并在平均性能上保持较大的差距。DMHOR, DMVH, SIDMH和FGCMH是基于深度学习的多模态哈希方法, 尽管这些方法具有很强的特征提取能力, 但是它们需要依赖大量的高质量数据去训练模型, 在数据量较小的Pascal Sentence数据集上, 其检索性能受到限制。如表4所示, 与上述深度方法相比, SPQH取得了最佳的检索精度。此外, 将本文提出的SPQH模型应用于仅含一种模态而其他模态缺失的单模态询问场景。'SPQH_img'表示询问编码阶段的样本仅提供图像模态, 'SPQH_txt'表示询问样本仅提供文本模态。'SPQH_img'和'SPQH_txt'首先分别利用学习到的重构矩阵 U_1 和 U_2 生成伪文本模态特征和伪图像模态特征, 然后通过习得的哈希函数获得其联合哈希码特征。图3(a)和图4(a)分别是完全未配对的图像询问场景和文本询问场景的PR曲线, 可以看出SPQH在图像检索和文本检索任务中的性能都优于其他的对比方法。

表2 完全配对的询问场景下不同比特长度的多模态检索任务的mAP比较

任务	方法	Pascal Sentence				NUS-WIDE				IAPR TC-12			
		16	32	64	128	16	32	64	128	16	32	64	128
	ITQ	0.3602	0.3523	0.3675	0.3803	0.3724	0.3751	0.3776	0.3789	0.3730	0.3844	0.3936	0.4020
	LSH	0.1011	0.1243	0.1572	0.2129	0.3421	0.3554	0.3544	0.3672	0.3251	0.3363	0.3509	0.3686
	DLLH	0.3631	0.3720	0.3971	0.3959	0.3738	0.3782	0.3794	0.3823	0.3644	0.3796	0.3863	0.3868
	HCOH	0.2135	0.4812	0.4846	0.4860	0.3232	0.3451	0.3434	0.3645	0.3082	0.3581	0.3717	0.3712
	MFH	0.1834	0.2399	0.2729	0.2731	0.3673	0.3752	0.3803	0.3815	0.3263	0.3374	0.3435	0.3451
	MVLH	0.1192	0.1347	0.1200	0.1202	0.3363	0.3339	0.3324	0.3284	0.3394	0.3401	0.3409	0.3499
O2O	OMH-DQ	0.4177	0.6719	0.7414	0.7622	0.5223	0.5381	0.5823	0.5957	0.3949	0.4200	0.4446	0.4642
	SIDMH	0.6681	0.7479	0.7596	0.7660	0.5828	0.5976	0.6055	0.6120	0.4131	0.4277	0.4364	0.4706
	AMFH	0.6837	0.7501	0.7511	0.7519	0.6190	0.6240	0.6271	0.6385	0.4198	0.4374	0.4571	0.4887
	SPQH_all	0.5176	0.7641	0.7799	0.7689	0.6208	0.6269	0.6353	0.6410	0.4060	0.4601	0.4691	0.4964

(2) NUS-WIDE数据集的对比实验：所有对比方法在NUS-WIDE数据集上的mAP值如表2所示，可以看出所提出的SPQH在不同码长设置下的“O2O”任务上都优于所有对比方法。图2(b)绘制所有方法的PR曲线，SPQH相比基准方法有显著的改进。表3记录了所有方法在I2I和T2T任务上的检索准确率,在平均检索准确率方面，本文提出的SPQH相比最佳基线方法STCH，在“I2I”任务上

高出4.36%，在“T2T”任务上获得了5.08%的性能提升。图3(b)和图4(b)分别描绘图像检索图像和文本检索文本的PR曲线，显然，SPQH的性能是最好的。上述实验结果表明本文提出的方法能够有效地完成图像搜索和文本搜索。通过观察表2、表3中SPQH_all, SPQH_img和SPQH_txt的mAP值，可以发现SPQH_txt比SPQH_img更接近SPQH_all。这种实验现象表明提出的模型在模态

表 3 完全未配对的询问场景下不同比特长度的多模态检索任务的mAP比较

任务	方法	Pascal Sentence				NUS-WIDE				IAPR TC-12			
		16	32	64	128	16	32	64	128	16	32	64	128
I2I	SPH	0.2425	0.2839	0.2919	0.3034	0.3142	0.3210	0.3465	0.3609	0.3352	0.3500	0.3634	0.3798
	SSMH	0.2600	0.3240	0.3499	0.3817	0.3468	0.3591	0.3877	0.4047	0.3258	0.3528	0.3503	0.3699
	UCMH	0.2766	0.3908	0.4778	0.5119	0.3500	0.3588	0.4051	0.4166	0.3487	0.3744	0.3795	0.3798
	STCH	0.2999	0.4126	0.4953	0.5648	0.3589	0.3603	0.3926	0.4260	0.3561	0.3807	0.3822	0.3912
	SPQH_img	0.3168	0.4533	0.5996	0.6111	0.3597	0.3612	0.4286	0.4319	0.3777	0.3823	0.3902	0.4003
T2T	SPH	0.3717	0.3958	0.4122	0.4534	0.4017	0.4128	0.4203	0.4399	0.2922	0.3021	0.3448	0.3485
	SSMH	0.3916	0.4443	0.4929	0.5341	0.5322	0.5538	0.5727	0.5894	0.3306	0.3529	0.3674	0.3879
	UCMH	0.4633	0.5324	0.5911	0.6317	0.5512	0.5677	0.5824	0.5878	0.3527	0.3742	0.3922	0.4157
	STCH	0.4718	0.6377	0.6651	0.6818	0.5677	0.6025	0.6148	0.6281	0.3789	0.3879	0.4206	0.4379
	SPQH_txt	0.5058	0.7433	0.7578	0.7600	0.5801	0.6199	0.6312	0.6327	0.3801	0.4054	0.4377	0.4416

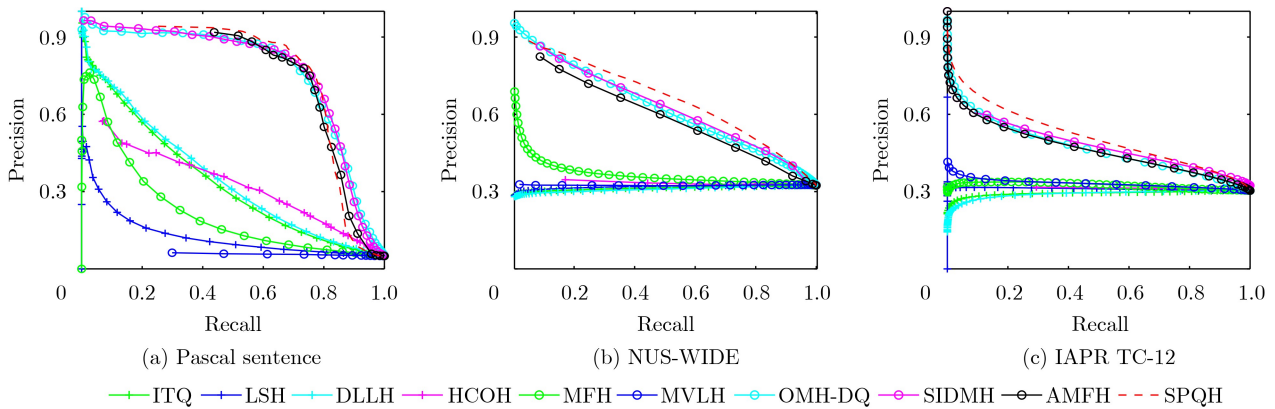


图 2 完全配对的询问场景下的PR曲线

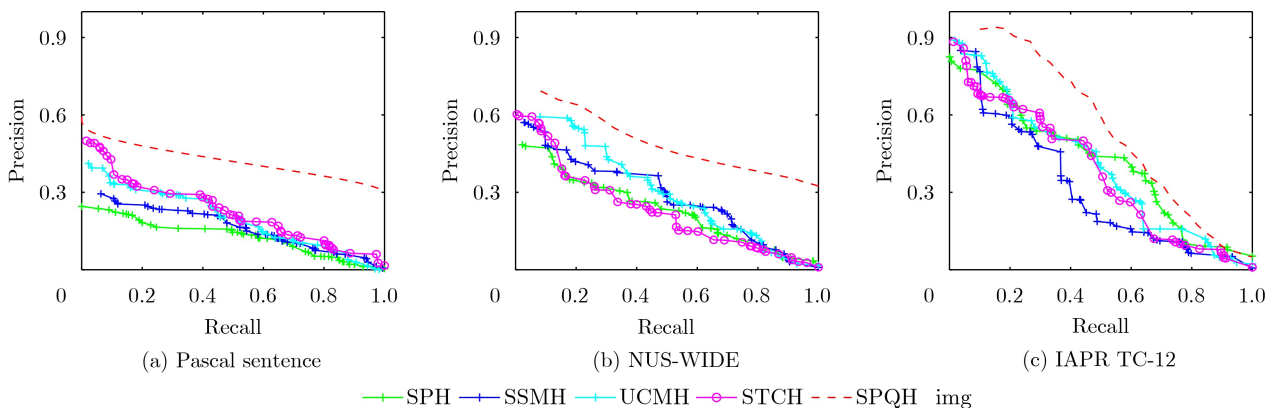


图 3 完全未配对的图像询问场景下的PR曲线

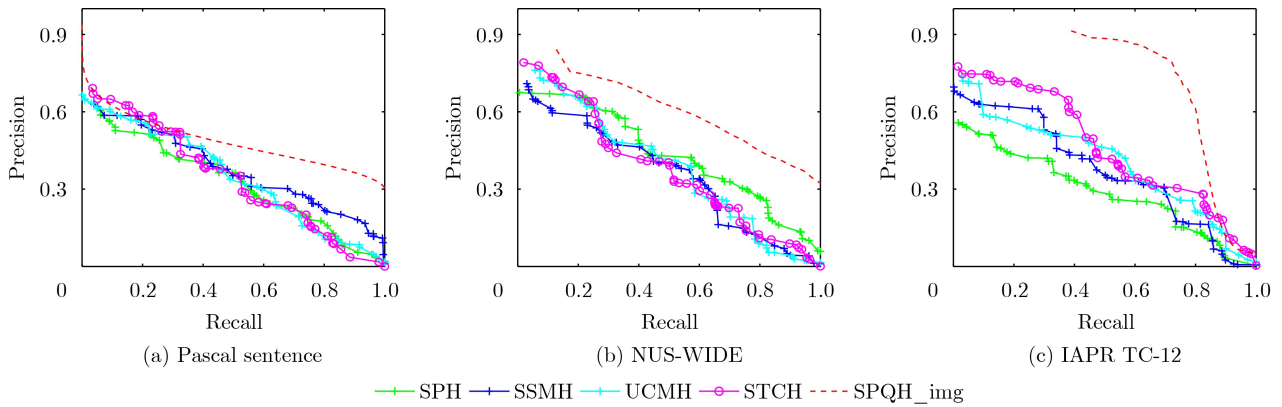


图4 完全未配对的文本询问场景下的PR曲线

表4 Pascal Sentence数据集上SPQH与深度方法的mAP比较

方法	哈希编码长度		
	32	64	128
DMHOR	0.5930	0.6693	0.6860
DMVH	0.5301	0.6010	0.6720
SIDMH	0.7479	0.7596	0.7660
FGCMH	0.7089	0.7433	0.7511
SPQH_all	0.7641	0.7799	0.7689

缺失的询问场景中通过已知文本模态生成图像模态特征的能力优于图像模态生成缺失的文本模态特征。

(3) IAPR TC-12数据集的对比实验：所有方法的比较结果及其在“O2O”任务上的PR曲线分别如表2和图2(c)所示。从实验结果中可以看到，本文提出的SPQH方法始终优于其他方法，并且随

着哈希码长度的增加，获得了更高的准确率。具体来说，本文提出的方法在“O2O”任务上获得了2.86%的提升。同样，在IARP TC-12数据集上，SPQH_img和SPQH_txt比最好的基线方法分别高出4.03%和3.95%。图3(c)和图4(c)分别描绘图像询问场景和文本询问场景下的PR曲线，实验结果表明所提出的SPQH方法在未成对的一种模态场景下的检索性能优于所有半配对的多模态哈希基线方法。

3.3 讨论

本节通过实验探究本文提出的方法在各特征空间的特征表征能力。以单标签的Pascal Sentence数据集为例开展直观的对比较验证实验。首先随机选取600个图像-文本样本来训练本文提出的模型，保存中间过程的多种特征。然后，使用t-SNE工具分别绘制这些特征的分布图。图5给出了原始图像特征、

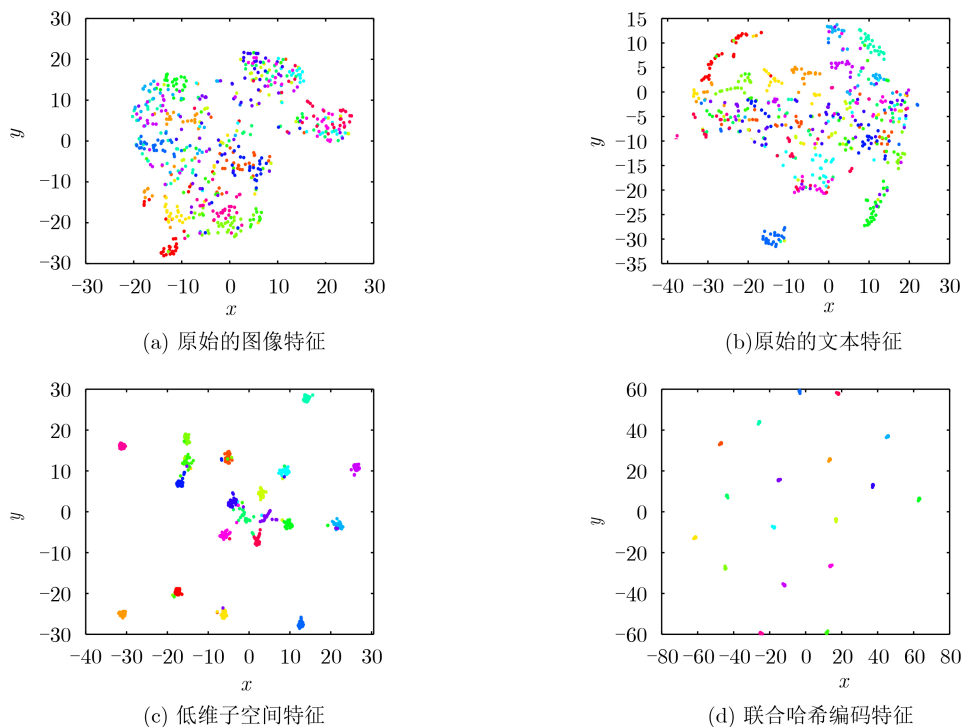


图5 不同特征空间的t-SNE可视化

原始文本特征、子空间特征和二值特征的可视化分布。如图5(a)和图5(b)所示，在原始模态空间中很难将各类数据分离。图5(c)中展示了文本模态的低维子空间的数据分布，可以观察到不同的聚类簇被呈现，不同类别的样本被聚集在不同的区域。图5(d)绘制的是学习到的联合哈希码特征的分布情况。在图5(d)中，可以清楚地看到具有相同类的样本更加紧凑，不同类别的区域更加分离。上述实验现象表明，本文提出的SPQH算法能够学习到具有判别性的联合哈希码特征。为了探索目标函数式(3)中的语义结构保持项 $\|S - B^T B\|_F^2$ 对模型性能的影响，本文设计了如下的消融实验，其中SPQH-DSP表示SPQH模型丢弃语义结构保持项。在3个数据集上的对比实验结果如表5所示。相比SPQH-DSP，SPQH保持了较大的检索性能优势。由此可见，本文构建的语义保持项是非常有助于判别性的哈希函数学习。提出的SPQH模型通过训练阶段习得的重构基矩阵对缺失的模态特征进行补全，然后再利用训练阶段习得的联合哈希函数生成联合哈希码以实现哈希检索。本节通过设计以下实验来探索SPQH模型的跨模态特征重构能力。SPQH_all是利用联合哈希函数对完全配对的图文询问样本进行联合哈希编码。SPQH1指将习得的联合哈希函数中的视觉模态的权重总是设置为0，这意味着在编码阶段仅对多模态询问样本的文本特征进行编码。SPQH2表示文本模态的权重则被设置为0，即仅针对图像特征进行哈希编码。SPQH_img和SPQH_txt首先利用了跨模态重构矩阵对缺失的模态特征进行补全，构成全配对的伪数据，然后再生成联合的哈希码。值得注意的是SPQH_all, SPQH_img,

SPQH_txt, SPQH1和SPQH2训练阶段没有区别，区别在于询问编码阶段。SPQH_all代表无缺失的询问集，而SPQH_img和SPQH2表示询问样本缺失文本，SPQH_txt和SPQH1表示询问样本缺失图像。如表6所示，SPQH_img和SPQH_txt的准确性分别高于SPQH1和SPQH2，检索结果表明本文提出的跨模态重构模块能够有效地生成缺失的模态特征。通过对比表3中记录的SPQH_img和SPQH_txt的mAP值，可以发现SPQH_img优于SPQH_img。造成这种实验现象的可能原因在于文本模态蕴含更多的语义信息，而视觉图像体现更丰富的细节信息，高层的语义更容易转换为视觉表达。因此，相比图像模态重构缺失的文本特征，提出的SPQH模型更善于利用已知的文本模态信息去重构缺失的图像特征。此外，在测试阶段，通过对询问集中的成对的多模态样例的占比进行不同设置以观察模型的性能变化。将配对样本占比分别设置为0%, 20%, 40%, 60%, 80%和100%，假设未成对的多模态样本都只提供了文本模态信息，实验中缺失的图像模态特征信息将被重构。图6绘制了3个数据集上不同的配对样本占比的实验结果。从实验结果中可以发现在视觉模态完全缺失的情况下仍然保持较高的性能，随着配对样本占比的增大，SPQH的性能提升非常细微。由此可见，SPQH模型能够

表 5 语义结构保持项的消融实验

方法	Pascal Sentence	NUS-WIDE	IAPR TC-12
SPQH-DSP	0.4614	0.4166	0.3949
SPQH	0.7799	0.6353	0.4691

表 6 SPQH在询问阶段基于不同模态特征编码的mAP比较

方法(询问数据的特征类型)	Pascal Sentence	NUS-WIDE	IAPR TC-12
SPQH1 (文本特征)	0.6159	0.4702	0.4162
SPQH_txt (文本特征+伪图片特征)	0.7578	0.6312	0.4377
SPQH2 (图片特征)	0.4837	0.4023	0.3680
SPQH_img (图片特征+伪文本特征)	0.5996	0.4286	0.3902
SPQH_all (图片特征+文本特征)	0.7799	0.6353	0.4691

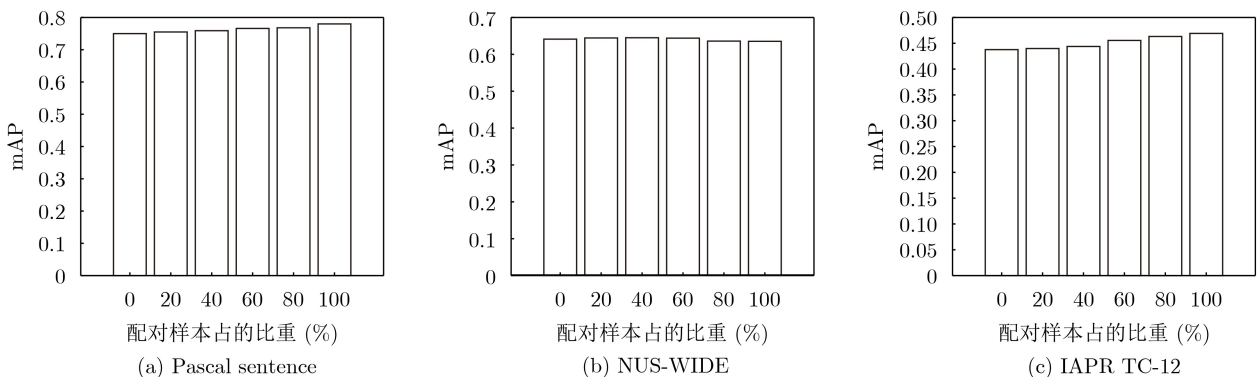


图 6 SPQH在询问编码阶段设置不同配对询问样本占比的mAP值

有效地利用文本模态对缺失的图片进行特征补全,并获得与完全成对的图文情景相接近的精度。通过上述讨论, SPQH模型能够学习到判别性的多模态联合哈希函数,且可以有效地对缺失地模态特征进行补全以实现半配对的询问场景下的联合哈希编码。

3.4 收敛性分析

在这一部分,从实验上来探究本文提出的模型的收敛性。当哈希码长度设置为128 bit时,图7给

出了本文的方法在3个数据集上的收敛曲线。其他码长的收敛曲线形状与128 bit码长下的收敛曲线形状类似。在图7中,每个子图的横轴表示迭代次数,纵轴表示目标函数值。通过观察收敛曲线的变化趋势,可以看出本文提出的算法在3个数据集上都快速下降并收敛到稳定值。尽管无法从理论上去证明所提算法的收敛性,但是通过上述的实验分析,本文的算法具备较好的收敛能力。

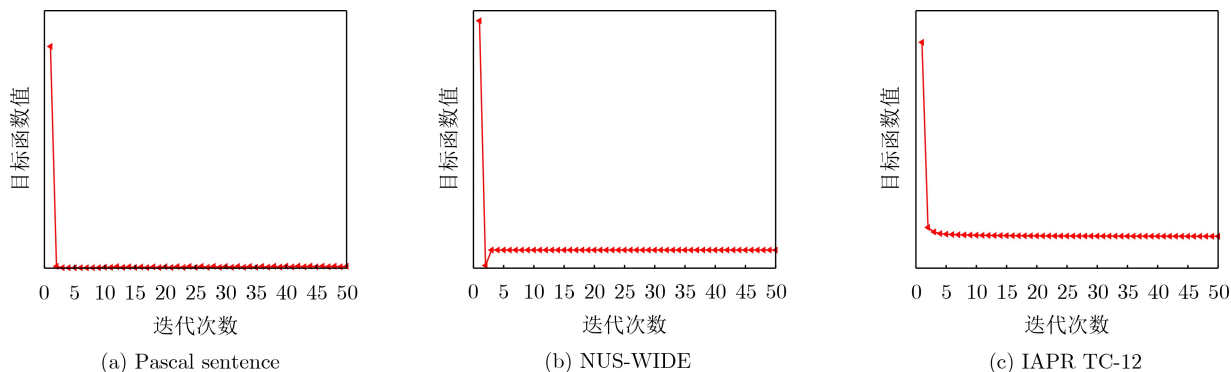


图7 本文方法在3个数据集上的收敛曲线

4 结束语

本文提出了一种面向多模态检索的半配对的多模态询问哈希编码方法,该方法保留了标签空间中的语义近邻结构信息并且有效地融合了多模态之间的互补信息以学习联合的多模态哈希函数。基于成对的多模态数据之间的语义一致特性,设计了一种跨模态重建学习模块,为未成对的多模态询问样本生成联合的哈希编码提供有效的解决思路。另外,提出一种快速的优化方法来求解目标函数,并在3个公共数据集上执行了广泛的检索实验,实验结果表明所提出的方法在半配对的多模态询问场景中具有优越的检索性能。算法存在以下不足之处:在编码阶段,对询问数据进行融合时的加权系数是固定的,没有考虑到询问样本间模态信息的差异性。在未来的工作中,将在深度网络框架中开发自适应的多模态询问编码策略,以进一步提高哈希模型的泛化性能。

参考文献

- [1] GEETHA V and SUJATHA N. A survey on divergent classification of social media networking[C]. 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022: 203–207. doi: [10.1109/ICCCIS56430.2022.10037606](https://doi.org/10.1109/ICCCIS56430.2022.10037606).
- [2] 顾广华, 霍文华, 苏明月, 等. 基于非对称监督深度离散哈希的图像检索[J]. 电子与信息学报, 2021, 43(12): 3530–3537. doi: [10.11999/JEIT200988](https://doi.org/10.11999/JEIT200988).
- [3] GONG Yunchao, LAZEBNIK S, GORDO A, et al. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval[J]. *IEEE Transactions 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] DATAR M, IMMORLICA N, INDYK P, et al. Locality-sensitive hashing scheme based on p-stable distributions[C]. The 20th Annual Symposium on Computational Geometry, Brooklyn, USA, 2004: 253–262. doi: [10.1145/997817.997857](https://doi.org/10.1145/997817.997857).
- [5] SHEN Fumin, SHEN Chunhua, LIU Wei, et al. Supervised discrete hashing[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015: 37–45. doi: [10.1109/CVPR.2015.7298598](https://doi.org/10.1109/CVPR.2015.7298598).
- [6] JI Rongrong, LIU Hong, CAO Liujuan, et al. Toward optimal manifold hashing via discrete locally linear embedding[J]. *IEEE Transactions on Image Processing*, 2017, 26(11): 5411–5420. doi: [10.1109/TIP.2017.2735184](https://doi.org/10.1109/TIP.2017.2735184).
- [7] KOUTAKI G, SHIRAI K, and AMBAI M. Hadamard coding for supervised discrete hashing[J]. *IEEE Transactions on Image Processing*, 2018, 27(11): 5378–5392. doi: [10.1109/TIP.2018.2855427](https://doi.org/10.1109/TIP.2018.2855427).
- [8] LIN Mingbao, JI Rongrong, LIU Hong, et al. Supervised online hashing via hadamard codebook learning[C]. The 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 2018: 1635–1643. doi: [10.1145/](https://doi.org/10.1145/)

GU Guanghua, HUO Wenhua, SU Mingyue, et al. Asymmetric supervised deep discrete hashing based image retrieval[J]. *Journal of Electronics & Information Technology*, 2021, 43(12): 3530–3537. doi: [10.11999/JEIT200988](https://doi.org/10.11999/JEIT200988).

- 3240508.3240519.
- [9] 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. doi: [10.1109/TIP.2020.2981879](https://doi.org/10.1109/TIP.2020.2981879).
- [10] JIN Lu, LI Zechao, PAN Yonghua, *et al.* Weakly-supervised image hashing through masked visual-semantic graph-based reasoning[C]. The 28th ACM International Conference on Multimedia, Seattle, USA, 2020: 916–924. doi: [10.1145/3394171.3414022](https://doi.org/10.1145/3394171.3414022).
- [11] LI Zechao, TANG Jinhui, ZHANG Liyan, *et al.* Weakly-supervised semantic guided hashing for social image retrieval[J]. *International Journal of Computer Vision*, 2020, 128(8/9): 2265–2278. doi: [10.1007/s11263-020-01331-0](https://doi.org/10.1007/s11263-020-01331-0).
- [12] SONG Jingkuan, YANG Yi, HUANG Zi, *et al.* Effective multiple feature hashing for large-scale near-duplicate video retrieval[J]. *IEEE Transactions on Multimedia*, 2013, 15(8): 1997–2008. doi: [10.1109/TMM.2013.2271746](https://doi.org/10.1109/TMM.2013.2271746).
- [13] SHEN Xiaobo, SHEN Fumin, SUN Quansen, *et al.* Multi-view latent hashing for efficient multimedia search[C]. The 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015: 831–834. doi: [10.1145/2733373.2806342](https://doi.org/10.1145/2733373.2806342).
- [14] LIU Li, YU Mengyang, and SHAO Ling. Multiview alignment hashing for efficient image search[J]. *IEEE Transactions on Image Processing*, 2015, 24(3): 956–966. doi: [10.1109/TIP.2015.2390975](https://doi.org/10.1109/TIP.2015.2390975).
- [15] LU Xu, LIU Li, NIE Liqiang, *et al.* Semantic-driven interpretable deep multi-modal hashing for large-scale multimedia retrieval[J]. *IEEE Transactions on Multimedia*, 2021, 23: 4541–4554. doi: [10.1109/TMM.2020.3044473](https://doi.org/10.1109/TMM.2020.3044473).
- [16] YU Jun, HUANG Wei, LI Zuhe, *et al.* Hadamard matrix-guided multi-modal hashing for multi-modal retrieval[J]. *Digital Signal Processing*, 2022, 130: 103743. doi: [10.1016/j.dsp.2022.103743](https://doi.org/10.1016/j.dsp.2022.103743).
- [17] 庾骏, 黄伟, 张晓波, 等. 基于松弛Hadamard矩阵的多模态融合哈希方法[J]. *电子学报*, 2022, 50(4): 909–920. doi: [10.12263/DZXB.20210760](https://doi.org/10.12263/DZXB.20210760).
- YU Jun, HUANG Wei, ZHANG Xiaobo, *et al.* Multimodal fusion hash learning method based on relaxed Hadamard matrix[J]. *Acta Electronica Sinica*, 2022, 50(4): 909–920. doi: [10.12263/DZXB.20210760](https://doi.org/10.12263/DZXB.20210760).
- [18] LU Xu, ZHU Lei, CHENG Zhiyong, *et al.* Online multi-modal hashing with dynamic query-adaption[C]. The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 2019: 715–724. doi: [10.1145/3331184.3331217](https://doi.org/10.1145/3331184.3331217).
- [19] YU Jun, ZHANG Donglin, SHU Zhenqiu, *et al.* Adaptive multi-modal fusion hashing via hadamard matrix[J]. *Applied Intelligence*, 2022, 52(15): 17170–17184. doi: [10.1007/s10489-022-03367-w](https://doi.org/10.1007/s10489-022-03367-w).
- [20] SHEN Xiaobo, SUN Quansen, and YUAN Yunhao. Semi-paired hashing for cross-view retrieval[J]. *Neurocomputing*, 2016, 213: 14–23. doi: [10.1016/j.neucom.2016.01.121](https://doi.org/10.1016/j.neucom.2016.01.121).
- [21] WANG Di, SHANG Bin, WANG Quan, *et al.* Semi-paired and semi-supervised multimodal hashing via cross-modality label propagation[J]. *Multimedia Tools and Applications*, 2019, 78(17): 24167–24185. doi: [10.1007/s11042-018-6858-8](https://doi.org/10.1007/s11042-018-6858-8).
- [22] GAO Jing, ZHANG Wenjun, ZHONG Fangming, *et al.* UCMH: Unpaired cross-modal hashing with matrix factorization[J]. *Neurocomputing*, 2020, 418: 178–190. doi: [10.1016/j.neucom.2020.08.029](https://doi.org/10.1016/j.neucom.2020.08.029).
- [23] JING Rongrong, TIAN Hu, ZHANG Xingwei, *et al.* Self-Training based semi-Supervised and semi-Paired hashing cross-modal retrieval[C]. 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022: 1–8. doi: [10.1109/IJCNN55064.2022.9892301](https://doi.org/10.1109/IJCNN55064.2022.9892301).
- [24] RASHTCHIAN C, YOUNG P, HODOSH M, *et al.* Collecting image annotations using amazon’s mechanical Turk[C]. The NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk, Los Angeles, America, 2010: 139–147.
- [25] CHUA T S, TANG Jinhui, HONG Richang, *et al.* NUS-WIDE: A real-world web image database from national university of Singapore[C]. The ACM International Conference on Image and Video Retrieval, Santorini, Greece, 2009: 48. doi: [10.1145/1646396.1646452](https://doi.org/10.1145/1646396.1646452).
- [26] ESCALANTE H J, HERNÁNDEZ C A, GONZALEZ J A, *et al.* The segmented and annotated IAPR TC-12 benchmark[J]. *Computer Vision and Image Understanding*, 2010, 114(4): 419–428. doi: [10.1016/j.cviu.2009.03.008](https://doi.org/10.1016/j.cviu.2009.03.008).
- [27] WANG Daixin, CUI Peng, OU Mingdong, *et al.* Deep multimodal hashing with orthogonal regularization[C]. The 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 2291–2297.
- [28] YANG Rui, SHI Yuliang, and XU Xinchun. Discrete multi-view hashing for effective image retrieval[C]. 2017 ACM on International Conference on Multimedia Retrieval, Bucharest, Romania, 2017: 175–183. doi: [10.1145/3078971.3078981](https://doi.org/10.1145/3078971.3078981).
- [29] LU Xu, ZHU Lei, LIU Li, *et al.* Graph convolutional multi-modal hashing for flexible multimedia retrieval[C/OL]. The 29th ACM International Conference on Multimedia, Chengdu, China, 2021: 1414–1422. doi: [10.1145/3474085.3475598](https://doi.org/10.1145/3474085.3475598).
- 庾 骏: 男, 博士, 讲师, 研究方向为多媒体信息搜索, 模式识别, 深度学习。
 马江涛: 男, 博士, 副教授, 研究方向为知识图谱, 深度学习。
 咸 阳: 男, 博士, 讲师, 研究方向为语音识别, 深度学习。
 侯瑞霞: 女, 博士, 副研究员, 研究方向为大数据分析, 数据挖掘。
 孙 伟: 男, 博士, 研究员, 研究方向为农林时空信息智能分析。

责任编辑: 马秀强