

基于松弛 Hadamard 矩阵的多模态融合哈希方法

庾 骏¹, 黄 伟¹, 张晓波¹, 尹贺峰²

(1. 郑州轻工业大学计算机与通信工程学院, 河南郑州 450000; 2. 江南大学计算机与人工智能学院, 江苏无锡 214000)

摘要: 哈希作为一种有效的数据表征技术, 已经在应对爆炸式增长的多媒体数据中扮演了重要的角色. 它由于低存储和高效率的优势, 在多媒体检索领域受到了越来越多的关注. 目前多模态哈希学习方法在多媒体检索任务中得到了较好的研究和发展. 然而, 多数的方法通过编码特征的内积重构对相似度来保持原始数据的结构信息, 但是带来较复杂的优化问题. 此外一些模型缺乏判别性使得检索性能的提升受到限制. 为了克服上述问题, 本文提出一种新型的多模态融合哈希方法, 在类别信息的监督下利用 Hadamard 矩阵为数据生成目标编码, 通过松弛严格的二值约束增大类间的间隔, 同时采用图嵌入的方式促进类内的紧凑性. 本文提出的方法既保证了模型具有很好的判别能力也简化了优化过程. 在3个公开数据集上的实验结果表明, 本文提出的方法在多媒体数据检索中是非常有效的, 平均性能上相比最优的对比方法提高了8.47%.

关键词: 哈希学习; 多模态融合; Hadamard 矩阵; 多媒体检索; 哈希中心

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2022)04-0909-12

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20210760

Multimodal Fusion Hash Learning Method Based on Relaxed Hadamard Matrix

YU Jun¹, HUANG Wei¹, ZHANG Xiao-bo¹, YIN He-feng²

(1. The College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan 450000, China;

2. The School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214000, China)

Abstract: Hashing, as an effective data representation technology, has played an important role in dealing with the explosive growth of multimedia data. Due to the advantages of its low storage and high efficiency, it has received more and more attention in the field of multimedia retrieval. At present, multi-modal hashing methods have been well researched and developed in multimedia retrieval tasks. However, most of these methods usually use the inner product of hashing features to reconstruct larger pairwise similarity, aiming to preserve the structural information of the original data, which will bring more complex optimization problems. Besides, some models lack discriminant ability, which leads to limitations in the improvement of retrieval performance. In order to overcome the above-mentioned problems, this paper proposes a new multimodal fusion hashing method. Under the supervision of category information, Hadamard matrix is used to generate target codes for data, and the margin between categories is increased by relaxing strict binary constraints. At the same time, the graph embedding approach is used to promote compactness within the class. The proposed method in this paper not only ensures the strong discriminative ability of the model, but also simplifies the optimization process. The experimental results on three public datasets show that the method proposed in this paper is very effective in multimedia data retrieval, and the average performance is 8.47% higher than that of the optimal comparison method.

Key words: hash learning; multimodal fusion; Hadamard matrix; multimedia retrieval; hash centers

1 引言

随着社交媒体和移动互联网的发展, 多媒体数据呈现爆炸式增长, 增大了信息搜索的挑战^[1]. 多媒体信

息通常以图像、文本、视频、音频等多种形式出现^[2]. 例如一个主题网页中包含了多种类型的媒体数据, 这些不同媒体数据间有很强的互补性. 哈希作为一种有效

的技术已经在信息搜索和模式识别领域引起了广泛的关注^[3]. 哈希方法编码高维的数据为二值编码,保留了原始数据间的相似性信息,这极大地加速了多媒体数据搜索的速度,同时节省了大量的存储空间^[4]. 目前,存在的哈希学习方法主要分为2种:单模态哈希^[5-11]和跨模态哈希^[13-18]. 众多方法中具有代表性的深度协同嵌入(Deep Collaborative Embedding, DCE)模型^[19]联合端对端学习和协同因子分析以学习图像和标签的公共潜在空间,并且同时解决跨模态搜索、基于内容的图像搜索和标签扩展问题. 不同于单模态哈希和跨模态哈希方法,多模态哈希联合多个模态数据综合地表征实例样本. 目标查询和数据库数据都由异构多模态特征描述. 每一种模态都从不同的角度来描述实例,并有自己的特点. 因此,有必要将不同模态特征结合起来,综合表示查询和数据库数据,以实现准确的检索. 一种简单的方式是将多个模态特征串联起来,作为单模态哈希模型的输入. 然而,这种拓展方式产生较多冗余信息,不能很好地挖掘多模态的互补性,而且可能会造成维度灾难. 基于此观察,一些多模态哈希方法被提出. 多特征哈希(Multiple Feature Hashing, MFH)^[20]同时保留每种模态数据的局部结构,并全局地考虑所有模态的局部结构来学习一组哈希函数,以映射媒体数据到低维的哈希编码. 多特征核哈希(Multiple Feature Kernel Hashing, MFKH)^[21]阐述多个模态数据的最优线性组合相似度. 多视图潜在哈希(Multi-view Latent Hashing, MVLH)^[22]旨在寻找一个统一的核特征空间,在此空间中不同模态数据被融合,然后通过多个模态共享的潜在因子学习二值编码. 多视图离散哈希(Multi-view Discrete Hashing, MVDH)^[23]联合执行矩阵分解和谱聚类以学习紧凑的哈希编码,这种联合学习方法使得 MVDH 产生更具判别性的哈希编码. 线上多模态哈希(Online Multimodal Hashing with Dynamic Query-adaption, OMH-DQ)^[24]在成对标签的监督下学习哈希编码. OMH-DQ 自适应地在线编码查询数据以捕获数据动态变化.

尽管上述方法已经取得了较好的进展,但是这些方法依然存在一些缺点:(1)大多数多模态哈希方法采用内积操作来保持同类样本编码一致性和异类样本间的差异性,由于二值化离散约束的限制,解决模型优化的问题存在较大的挑战;(2)采用固定的权值对多模态数据进行哈希编码,在复杂的多模态数据场景中不能捕捉多模态数据之间的互补信息,导致检索精度提升有限. 本文中提出了一种新的多模态哈希方法(Multimodal Fusion Hashing based on Relaxed Hadamard matrix, MFH-RH)有效地解决了上述问题. 图1是MFH-RH的框架示意图. MFH-RH采用了两阶段的学习方式保证了学习的哈希编码具有较强的判别能力,在优化过

程得到了离散变量的闭合解,提高了优化的效率,同时以一种自适应的学习方式有效地捕捉模态间的互补信息,增强了模型在复杂数据环境中的适应能力和鲁棒性. 代表性的哈希方法和本文提出的方法的主要特点被总结在表1中. 本文提出的MFH-RH框架的创新点如下.

(1) MFH-RH引入Hadamard矩阵,在类别信息的监督下为多模态数据生成目标编码,促使同类数据逼近相同的哈希中心. 进一步地,对目标编码进行了松弛,增大了类间的间隔,同时采用图嵌入方法保证类内的紧凑型,因而增强了学习到的投影矩阵的判别能力.

(2) MFH-RH采用两阶段的学习方式,能够自适应地捕捉样本模态间的互补信息,增强模型在复杂多模态数据场景中的哈希编码能力. 同时在优化过程中以闭合解的形式求解离散变量,有效地改善了优化求解效率,缓解了因二值化离散约束带来的挑战.

表1 代表性哈希方法和本文提出方法的主要特点

方法	方法类型	学习方式	图嵌入方法	引入Hadamard矩阵	两阶段学习
DLLE	单模态哈希	无监督	是	否	否
HCOH	单模态哈希	有监督	否	是	否
MFH	多模态哈希	无监督	是	否	否
MFKH	多模态哈希	有监督	是	否	否
MVLH	多模态哈希	无监督	否	否	否
MVDH	多模态哈希	无监督	是	否	否
OMH-DQ	多模态哈希	有监督	否	否	是
MFH-RH	多模态哈希	有监督	是	是	是

2 算法模型

2.1 问题描述

假定训练数据包含 n 个多模态实例样本,每个实例由 M 个不同媒体数据组成(如图片和文本),这些媒体数据是成对的且语义内容相似. 训练集中的第 M 媒体数据表示为 $\mathbf{X}_m = [\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_n^{(m)}] \in \mathbf{R}^{d_m \times n}$,其中 d_m 是第 M 媒体数据的维度. 在训练集中,给定同类实例样本的类别信息. r 表示实例样本的哈希编码长度,则第 i 个实例样本可表示成 $\mathbf{b}_i = \{+1, -1\}^r$. 本文方法的目的是根据已知的数据信息学习鉴别性的投影函数. 利用训练阶段学习的投影函数自适应地编码新到达的多媒体数据,生成可靠的哈希码. 在本文方法中,预先定义了每个类样本的哈希中心,促使相同类别的样本在汉明空间中逼近其类中心,不同类样本趋近于不同的类别哈希中心. $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\} \in \mathbf{R}^{r \times k}$ 分别表示 k 个类别数据的哈希中心,其中 r 表示哈希中心的维度. 在汉明空间选择合适的哈希中心需要满足如下要求:不同类别的哈希中心点之间的间距应该确保足够大,以使不同类样本在汉明空间中能够很好地分离开. 定义一个有效的哈

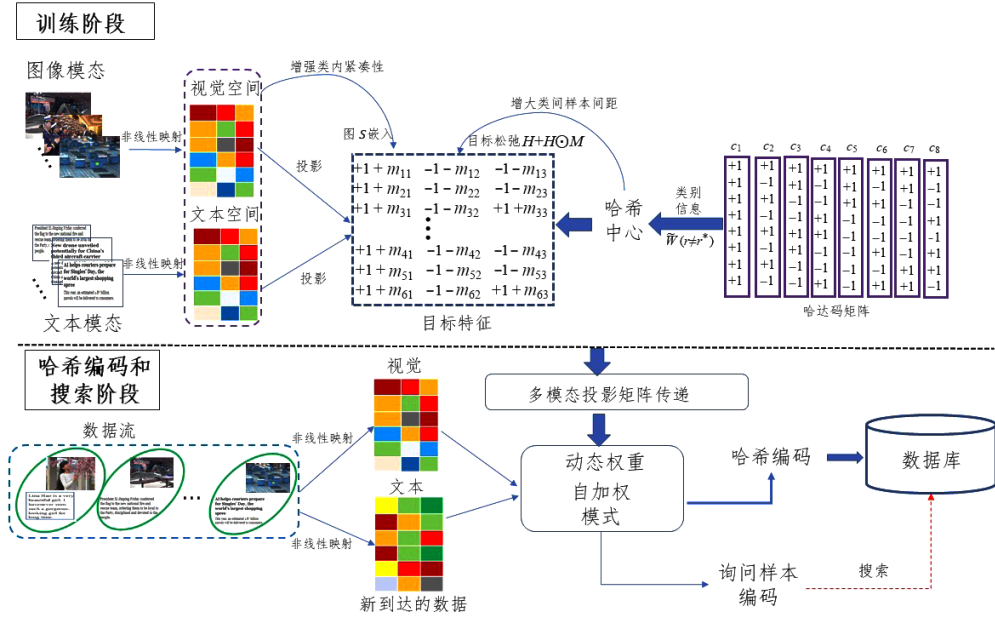


图1 本文中MFH-RH的框架示意图

注:提出的模型分为2个阶段。(1)训练阶段:该阶段引入了Hadamard矩阵生成训练数据的目标编码,投影后的同类数据逼近它们的哈希中心。为了增强类内的紧凑性和类间的松散性,在投影矩阵学习过程中分别增加图正则项和松散项。(2)哈希编码阶段:为了捕捉样本中不同媒体数据的差异性,采用一种自适应的加权方案编码批数据,生成的哈希编码存储在数据库中,对任意的询问样本,首先为其生成哈希特征,然后与数据库中样本的编码匹配,返回最近邻的样本作为搜索结果。

哈希中心集,如定义一所示。

定义1 哈希中心 $C = \{c_i\}_{i=1}^s \subset \{+1, -1\}^r$ 满足所有两两中心点间的距离的平均值大于或者等于 $r^*/2$, 即

$$\frac{1}{N} \sum_{i \neq j}^k D_H(c_i, c_j) \geq \frac{r^*}{2} \quad (1)$$

其中, N 是两两哈希中心的组合数目, k 是类别数目, D_H 表示汉明距离。

通过 Sylvester 方法^[25]生成的 Hadamard 矩阵具有以下属性:

(1) 它是一个 r^* 阶的方阵, 其中的元素是 +1 或者 -1. r^* 满足

$$r^* = \min \{l | l = 2^{\text{num}}, l \geq r, l \geq k, \text{num} = 1, 2, \dots\} \quad (2)$$

(2) 它的每行之间和每列之间都是正交的, 这就确保了任意 2 列或 2 行之间的汉明距离为 $r^*/2$ 。

通过上述分析, 可以看出 Hadamard 矩阵的列向量或者行向量满足定义 1, 可以作为类哈希中心点。值得注意的是生成的哈希中心的维数 r^* 不一定等于输出编码长度 r 。为了解决这个问题, 采用局部敏感哈希 (Local Sensitive Hashing, LSH)^[26] 来转换 Hadamard 矩阵, 以确保哈希中心和输出编码的长度一致。具体为

$$\tilde{c}_i = \text{sign}(\tilde{W}^T c_i) \quad (3)$$

其中, $\tilde{W} = \{\tilde{w}_i\}_{i=1}^r \in \mathbf{R}^{r^* \times r}$ 从标准的高斯分布中抽样获得。转换后的 Hadamard 矩阵保留了原始 Hadamard 矩阵的主要属性信息, 同时符合列之间最小汉明距离的

要求。详细的理论证明可参考文献[27]。此外, 针对归属多类的多标签样本, 其对应的哈希中心是多个特定类中心 ($C = \{c_1, c_2, \dots, c_k\}$ 分别对应 k 个种类的哈希中心) 的质点。通过上述过程, 获得最终训练数据的目标输出编码 $H \in \mathbf{R}^{r \times n}$ 。为了保证类内的紧凑和类间的分离, 利用松弛后的 Hadamard 矩阵为类数据分配类中心编码, 同时采用图正则化保持类内数据的邻接关系。生成的哈希特征自适应地捕捉多模态数据间的互补信息以提高在复杂的数据场景中的编码质量, 进而提升检索精度。总的框架包含个阶段: 训练阶段和哈希编码阶段。

2.2 训练阶段

针对第 m 媒体数据 X_m , 首先计算它的非线性表示 ϕ_m , 其中 ϕ_m 中的第 i 列为 $\phi_m^i = \left[\exp\left(\frac{\|x_i^{(m)} - a_1^{(m)}\|_F^2}{2\sigma_m^2}\right), \dots, \exp\left(\frac{\|x_i^{(m)} - a_p^{(m)}\|_F^2}{2\sigma_m^2}\right) \right]$, $\{a_j^{(m)}\}_{j=1}^p$ 是从第 m 媒体数据中采用 k-means 聚类算法计算获得的 p 个锚点样本, σ_m 表示高斯核参数。可以获得如下映射模型:

$$\begin{aligned} \min_{\mu_m, W_m} \sum_{m=1}^M (\mu_m)^y \|H - W_m \phi_m\|_F^2 + \lambda \sum_{m=1}^M \|W_m\|_F^2 \\ \text{s.t.} \sum_{m=1}^M \mu_m = 1, \mu_m \geq 0 \end{aligned} \quad (4)$$

其中, μ_m 是第 m 媒体数据对应的权值系数, W_m 是第 m 媒体数据的投影矩阵。

H 被严格地限制为二值编码, 为了使其具有更大

的自由度,将严格的二元约束放宽为软约束.使用松弛变量矩阵 $\mathbf{E} = \mathbf{H} + \mathbf{H} \odot \mathbf{Q}$ 来取代目标编码矩阵 \mathbf{H} ,其中 \odot 是矩阵的 Hadamard 乘积,具体的推导过程见附录 1. 式(4)能被改写成

$$\begin{aligned} \min_{\mu_m, \mathbf{W}_m, \mathbf{V}, \mathbf{Q}} & \sum_{m=1}^M (\mu_m)^\gamma \| \mathbf{V} - \mathbf{W}_m \boldsymbol{\phi}_m \|_{\text{F}}^2 + \beta \| \mathbf{V} - (\mathbf{H} + \mathbf{H} \odot \mathbf{Q}) \|_{\text{F}}^2 + \\ & \lambda \sum_{m=1}^M \| \mathbf{W}_m \|_{\text{F}}^2 \\ \text{s.t.} & \sum_{m=1}^M \mu_m = 1, \mu_m \geq 0, \mathbf{Q} \geq 0 \end{aligned} \quad (5)$$

其中, \mathbf{V} 是低维子空间的表示矩阵, β 和 λ 表示惩罚参数.

为了促进类内的紧凑型,构造类紧凑型图.邻接图被定义如下:

$$\mathbf{S}_{ij} = \begin{cases} \exp\left(-\frac{\|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\|^2}{\sigma}\right), & \text{如果 } i \text{ 和 } j \text{ 是同类样本} \\ 0, & \text{否则} \end{cases} \quad (6)$$

其中, σ 是热核参数, $\boldsymbol{\phi}_i = [\boldsymbol{\phi}_1^i; \boldsymbol{\phi}_2^i; \dots; \boldsymbol{\phi}_M^i]$ 是多个配对的不同媒体数据的级联特征.在转换后的子空间中,同类样本应该保持紧密.基于图的正则项如下所示:

$$\min \| \mathbf{v}_i - \mathbf{v}_j \|^2 \mathbf{S}_{ij} = \text{Tr}(\mathbf{V} \mathbf{L} \mathbf{V}^T) \quad (7)$$

其中, $\mathbf{L} = \mathbf{D} - \mathbf{S}$; \mathbf{D} 是一个对角矩阵,它的第 i 个元素定义为 $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{i,j}$; $\text{Tr}(\cdot)$ 表示矩阵的迹.

联合式(5)和式(7),总的目标函数如下所示:

$$\begin{aligned} \min_{\mu_m, \mathbf{W}_m, \mathbf{V}, \mathbf{Q}} & \sum_{m=1}^M (\mu_m)^\gamma \| \mathbf{V} - \mathbf{W}_m \boldsymbol{\phi}_m \|_{\text{F}}^2 + \alpha \text{Tr}(\mathbf{V} \mathbf{L} \mathbf{V}^T) + \beta \| \mathbf{V} - \\ & (\mathbf{H} + \mathbf{F} \odot \mathbf{Q}) \|_{\text{F}}^2 + \lambda \sum_{m=1}^M \| \mathbf{W}_m \|_{\text{F}}^2 \\ \text{s.t.} & \sum_{m=1}^M \mu_m = 1, \mu_m \geq 0, \mathbf{Q} \geq 0 \end{aligned} \quad (8)$$

其中, α, β 和 λ 是可调参数.

上述目标函数(8)是非凸问题,很难直接对其优化求解.采用常用的交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)对每个变量进行交替优化.具体的过程如下.

步骤 1: 固定其他变量,优化 \mathbf{W}_m . 很容易获得关于 \mathbf{W}_m 的子问题,如下所示:

$$\min_{\mathbf{W}_m} \sum_{m=1}^M (\mu_m)^\gamma \| \mathbf{V} - \mathbf{W}_m \boldsymbol{\phi}_m \|_{\text{F}}^2 + \lambda \sum_{m=1}^M \| \mathbf{W}_m \|_{\text{F}}^2 \quad (9)$$

令式(9)关于 \mathbf{W}_m 的偏导等于零,可以获得其解:

$$\mathbf{W}_m = (\mu_m)^\gamma \mathbf{V} \boldsymbol{\phi}_m^T ((\mu_m)^\gamma \boldsymbol{\phi}_m \boldsymbol{\phi}_m^T + \lambda \mathbf{I})^{-1} \quad (10)$$

步骤 2: 固定其他变量,优化 \mathbf{V} . 假设 $\mathbf{E} = \mathbf{H} + \mathbf{H} \odot \mathbf{Q}$, $\mathbf{Q} \geq 0$, 其子问题如下:

$$\min_{\mathbf{V}} \sum_{m=1}^M (\mu_m)^\gamma \| \mathbf{V} - \mathbf{W}_m \boldsymbol{\phi}_m \|_{\text{F}}^2 + \alpha \text{Tr}(\mathbf{V} \mathbf{L} \mathbf{V}^T) + \beta \| \mathbf{V} - \mathbf{E} \|_{\text{F}}^2 \quad (11)$$

同样令式(11)关于 \mathbf{V} 的偏导等于零,可得它的解为

$$\mathbf{V} = \sum_{m=1}^M (\mu_m)^\gamma \mathbf{W}_m \boldsymbol{\phi}_m + \beta \mathbf{E} \left(\sum_{m=1}^M (\mu_m)^\gamma \mathbf{I} + \beta \mathbf{I} + \alpha \mathbf{L} \right)^{-1} \quad (12)$$

步骤 3: 固定其他变量,优化 \mathbf{Q} . 令 $\mathbf{R} = \mathbf{V} - \mathbf{H}$, 则关于 \mathbf{Q} 的子问题如下:

$$\min_{\mathbf{Q}} \| \mathbf{R} - \mathbf{H} \odot \mathbf{Q} \|_{\text{F}}^2 \quad \text{s.t. } \mathbf{Q} \geq 0 \quad (13)$$

根据文献[28],能获得最优解如下:

$$\mathbf{Q} = \max(\mathbf{H} \odot \mathbf{R}, 0) \quad (14)$$

步骤 4: 固定其他变量,优化 μ_m . 保留与 μ_m 有关的项,可以获得

$$\begin{aligned} \min_{\mu_m \geq 0} & \sum_{m=1}^M (\mu_m)^\gamma G_m \\ \text{s.t.} & \sum_{m=1}^M \mu_m = 1 \end{aligned} \quad (15)$$

其中 $G_m = \| \mathbf{V} - \mathbf{W}_m \boldsymbol{\phi}_m \|_{\text{F}}^2$. 利用拉格朗日乘子法将式(15)转换成

$$\min_{\mu_m} \sum_{m=1}^M (\mu_m)^\gamma G_m + \zeta \left(1 - \sum_{m=1}^M \mu_m \right) \quad (16)$$

令式(16)关于 μ_m 的偏导等于零,获得最终的解为

$$\mu_m = \frac{(\gamma G_m)^{1/(1-\gamma)}}{\sum_{m=1}^M (\gamma G_m)^{1/(1-\gamma)}} \quad (17)$$

重复上述 4 个步骤,直到算法收敛.最后获得最优的投影矩阵 \mathbf{W}_m 和权重系数 μ_m . 哈希编码函数能被定义为

$f(\mathbf{x}_m^q) = \text{sign} \left(\sum_{m=1}^M (\mu_m)^\gamma \mathbf{W}_m \boldsymbol{\phi}_m (\mathbf{x}_m^q) \right)$, 其中 \mathbf{x}_m^q 表示询问样本的第 m 媒体数据.

2.3 哈希编码阶段

然而,从 2.2 节中学习获得的权重系数 μ_m 并不能很好地捕捉到动态数据的差异性.期望在编码较大差异的数据时,不同媒体(模态)对应的权值能够自适应地调整.受此思想的启发,采用一种自加权模式为新达到的多媒体数据进行哈希编码,以获得更加精确的哈希码.在编码阶段,假定数据是以批数据形式呈现.新达到的批数据被编码后存储在数据库中.自适应的哈希编码过程如下所示.

$$\begin{aligned} \min_{\mathbf{B}_q, \mu_m^q} & \sum_{m=1}^M (\mu_m^q)^\gamma \| \mathbf{B}_q - \mathbf{W}_m \boldsymbol{\phi}_m^q \|_{\text{F}}^2 \\ \text{s.t.} & \sum_{m=1}^M \mu_m^q = 1, \mathbf{B}_q \in \{-1, +1\}^{r \times n_q} \end{aligned} \quad (18)$$

其中, \mathbf{B}_q 和 n_q 分别是新数据实例样本的编码和数目, $\boldsymbol{\phi}_m^q$ 是新样本的非线性转换后的表示矩阵.采用交替更新 \mathbf{B}_q, μ_m^q 来获得数据最优的编码特征.具体如下.

固定 \mathbf{B}_q 更新 μ_m^q : 关于 μ_m^q 的最优值为

$$\mu_m^q = \frac{(\gamma \mathbf{G}_m^q)^{1/(1-\gamma)}}{\sum_{m=1}^M (\gamma \mathbf{G}_m^q)^{1/(1-\gamma)}} \quad (19)$$

其中, $\mathbf{G}_m^q = \|\mathbf{B}_q - \mathbf{W}_m \phi_m^q\|_F^2$.

固定 μ_m^q 更新 \mathbf{B}_q , 能获得其闭合解:

$$\mathbf{B}_q = \text{sign}\left(\sum_{m=1}^M (\mu_m^q)^\gamma \mathbf{W}_m \phi_m^q\right) \quad (20)$$

最后输出的 \mathbf{B}_q 被当成是新达到的批数据的最后输出编码, 并存储于数据库中以供询问数据查询. 本文方法的整个过程被总结在算法 1 中. 算法的复杂度分析如下: 计算 ϕ_m 需要 $\mathcal{O}(np)$, 则 M 个模态的计算复杂度为 $\mathcal{O}(Mnp)$; 更新 \mathbf{W}_m 耗费 $\mathcal{O}(np^2)$, 则 M 个模态的计算复杂度为 $\mathcal{O}(Mnp^2)$; 由于式 (15) 中的求逆过程只需计算一次, 因此计算 \mathbf{V} 的时间复杂度为 $\mathcal{O}(rn^2)$; 算法总的复杂度为 $\mathcal{O}(\text{iter} \times (Mnp^2 + rn^2))$, 其中 iter 是总的迭代次数.

算法 1 基于松弛 Hadamard 矩阵的多模态融合哈希 (MFH-RH)

训练阶段:

输入: 训练数据矩阵 $X_m \in \mathbf{R}^{d_m \times n}$, 数据的类别信息, 参数 α, β, γ 和 λ .

根据 2.1 节生成训练数据的目标输出编码 $\mathbf{H} \in \mathbf{R}^{r \times n}$.

计算训练数据的非线性转换矩阵 ϕ_m .

计算邻接矩阵 \mathbf{S} , 并得到图拉普拉斯矩阵 \mathbf{L} .

初始化: 投影矩阵 $\mathbf{W}_m, \mu_m, \mathbf{V}$

重复:

步骤 1 根据式 (10) 更新 \mathbf{W}_m ;

步骤 2 根据式 (12) 更新 \mathbf{V} ;

步骤 3 根据式 (14) 更新 \mathbf{Q} ;

步骤 4 根据式 (17) 更新 μ_m ;

直至收敛.

输出: 投影矩阵 $\mathbf{W}_m (m=1, \dots, M)$.

哈希编码阶段:

For $q = 1, \dots, T$ do

获取新达到的多媒体数据实例 \mathbf{X}_q

计算非线性表示矩阵 ϕ_m^q

重复:

根据式 (19) 计算 μ_m^q

根据式 (20) 计算 \mathbf{B}_q

直至收敛.

输出 \mathbf{B}_q 并存储在数据库

end

3 实验及结果分析

在这部分, 在 2 个公开数据集上执行多媒体检索实验来验证提出的 MFH-RH 的有效性. 实验数据、对比方法、评估方法和详细的实验结果将被详细介绍. 所有实验运行在 Intel (R) Core (TM) I7-10700 CPU @2.9GHz 和 16GB 内存的工作站上.

3.1 数据集

3 个常见的多模态数据集被用来评估所有对比算法和 MFH-RH 的检索性能.

Wiki 数据集^[29]是一个单标签多模态数据集, 总共包含 2866 个多媒体文件. 每个多媒体文件包含一张图片和一段文字, 且被划分为 10 个类别中的一类. 每张图片表示成 128 维 SIFT 直方图向量. 每段文字通过 Latent Dirichlet Allocation (LDA)^[30] 表示为一个 10 维的特征向量. 随机选取一个包含 2173 个文件的子集作为训练集和数据库, 剩余的 963 个文件作为询问测试集.

Pascal Sentence 数据集^[31]包含 1000 对图像-文本样本. 该数据集被划分为 20 个种类. 每个类别包含 50 对图像-文本. 随机从每类中抽选 30 对样本组成了包含 600 对样本的训练集. 剩余的样本作为询问测试集. 每张图片由 4096 维 CNN 视觉特征表示^[32]. 针对文本, 首先获得 300 维的 BoW (Bag of Word) 特征, 然后利用 LDA 模型获取在 100 个主题中的概率分布.

MirFlickr 数据集^[17]从 Flickr 网站收集而来. 数据集被划分为 24 个种类, 包含了 25 000 对图像-文本. 保留了至少出现 20 次的标签, 然后移除那些没有标签的数据. 每对样本被分配到多个类别. 因此, MirFlickr 是个多标签数据集. 每张图片被表示为 150 维的边缘直方图特征. 至于文本, 利用 Principal Component Analysis (PCA), 从 BoW 特征中抽取的 500 维向量作为其特征表示. 随机从包含 15 902 对样本的数据库中随机选择 5000 对数据作为训练集, 剩余 693 对图像文本作为询问测试集.

3.2 实验设置

在训练阶段, Hadamard 矩阵的列作为对应类的哈希中心. 对于单标签数据集 Wiki 和 Pascal Sentence, 数据的目标编码设置为它所属类的哈希中心. 而针对多标签数据集 MirFlickr, 将多个类哈希中心的质点作为数据的目标编码. 在实验中, 对比 7 种相关的哈希方法来验证提出的方法在多媒体检索任务中的有效性. 它们被划分为两大类. 一是多模态哈希方法, 包括 MFH^[20], MVLH^[22] 和 OMH-DQ^[24]. 二是单模态哈希方法, 包括 ITQ^[6], LSH^[7], DLLE^[8], HCOH^[9]. 由于单模态哈希方法并不能同时处理多个模态数据, 为了公平对比, 采用级联操作将多模态级联后的特征作为模型的输入. 参考原始论文中给出的参数候选范围, 对各个模型中的手动参数进行细心调整并记录最好的结果. 在提出的模型中, α 和 β 的值在一个较大候选范围 $\{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}, 1e^1\}$ 内进行调整. λ 在 Wiki, Pascal Sentence 和 MirFlickr 分别设置为 $1e^{-5}, 1e^{-2}$ 和 $1e^{-1}$. γ 设置为 0.5. 训练阶段, 算法的最大迭代次数为 50. 采用 k-means 聚类算法获得锚点样本, 在 3 个数据集上锚点数目 p 分别为 1000, 600, 1000. 使用平均精确度 (mean Aver-

age Precision, mAP)作为检索性能的评价指标. 针对一个询问数据 q , Average Precision(AP)的定义如下:

$$AP(q) = \frac{1}{l_q} \sum_{i=1}^R P_q(i) \delta_q(i) \quad (21)$$

其中, $P_q(i)$ 表示前 i 个检索结果的精度; 如果第 i 个结果是真的邻域样本, 则 $\delta_q(i) = 1$, 否则 $\delta_q(i) = 0$; l_q 是前 R 个检索结果的正确结果的数目; mAP 被定义成所有询问样本的平均 AP. 在本文实验中 R 设为数据库的尺度.

此外, 另一种广泛使用的评价标准——哈希查找(hash lookup)被采纳在对比实验中. 构造查找表以返回落在查询点的汉明半径 d 内的那些点. 计算所有查询的平均召回率和准确率. 根据哈希查找协议, 随着汉明半径 d 的增加, 绘制了 Precision-Recall(PR) 曲线以便更直观地评估性能.

3.3 对比实验结果

在这部分, 将讨论和分析在 3 个不同数据集上的实验结果. 详细的实验观察描述如下(表中加粗的数据表示性能最好的方法). 在 Wiki, Pascal Sentence 和 MirFlickr 数据集上的对比实验结果如表 2 所示. 可以清楚地观察到本文提出的 MFH-RH 在 Wiki 和 Pascal Sentence 上超过了所有的对比方法. 在 MirFlickr 上, 当哈希编码设置为 16bit 时, 本文提出的 MFH-RH 低于 OMH-DQ. 但是随着编码长度的增大, 本文提出的方法

在精度上高于所有方法. 本文提出的 MFH-RH 优于对比方法 HCOH. 具体地, 在 Wiki, Pascal Sentence 和 MirFlickr 上, MFH-RH 针对不同哈希编码长度设置下的检索精度的平均值分别高于 HCOH 14.7%, 34.1% 和 2.8%. HCOH 是一种基于 Hadamard 矩阵的线上哈希学习方法. 与 HCOH 对比的实验结果表明, 多模态数据的有效融合相比简单的级联特征在哈希学习中能更好地挖掘多模态数据间的互补信息. 本文方法在 Pascal Sentence 上提升相当明显的可能原因在于该数据集采用高维度的 CNN 特征, 级联后的多模态特征携带的冗余信息太多. 与 OMH-DQ 相比, 本文提出的方法在 Wiki 和 Pascal Sentence 上分别取得了平均 26% 和 10.9% 的精度提升. 这说明利用 Hadamard 矩阵引导哈希函数学习是非常有效的, 极大增强了哈希编码的判别性. 随着编码长度的增大, 本文提出的 MFH-RH 在性能上得到稍微改善. 在 Pascal Sentence 数据集上, 当哈希编码设置为 32 bit, 64 bit 和 128 bit 时, 检索性能没有获得进一步改善. 这说明本文方法对哈希编码长度不是特别敏感, 在较短编码长度设置下能够取得满意的检索精度. 图 2 展现了所有方法在 Wiki, Pascal Sentence 和 MirFlickr 上的 PR 曲线. 从图 2 中可以观察到, 本文提出的方法优越于其他的方法. 这些实验结果表明, 基于 Hadamard 矩阵的多模态融合哈希在多媒体数据检索中是非常有效的.

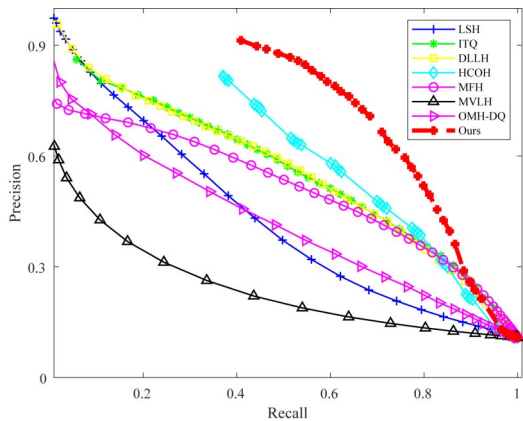
表 2 各算法在 Wiki, Pascal Sentence, MirFlickr 数据集上执行多媒体数据检索的 mAP 性能比较

算法	Wiki				Pascal Sentence				MirFlickr			
	16 bit	32 bit	64 bit	128 bit	16 bit	32 bit	64 bit	128 bit	16 bit	32 bit	64 bit	128 bit
ITQ	0.5122	0.5359	0.5490	0.5532	0.3602	0.3523	0.3675	0.3803	0.5821	0.5829	0.5854	0.5878
LSH	0.4306	0.4712	0.5085	0.5276	0.1011	0.1243	0.1572	0.2129	0.5515	0.5503	0.5494	0.5466
DLLE	0.5234	0.5330	0.5466	0.5506	0.3631	0.3720	0.3971	0.3959	0.5426	0.5432	0.5424	0.5426
HCOH	0.5450	0.5474	0.5494	0.5490	0.2135	0.4812	0.4846	0.4860	0.5921	0.6659	0.6704	0.6762
MFH	0.4630	0.5040	0.5455	0.5569	0.1834	0.2399	0.2729	0.2731	0.5818	0.5859	0.5848	0.5890
MVLH	0.3027	0.3166	0.3000	0.3045	0.1192	0.1347	0.1200	0.1202	0.5678	0.5701	0.5709	0.5682
OMH-DQ	0.4117	0.4393	0.4556	0.4319	0.4177	0.6719	0.7414	0.7662	0.6532	0.6767	0.6854	0.7030
MFH-RH	0.6890	0.6927	0.6980	0.7005	0.6867	0.7820	0.7820	0.7820	0.6251	0.6900	0.6950	0.7063

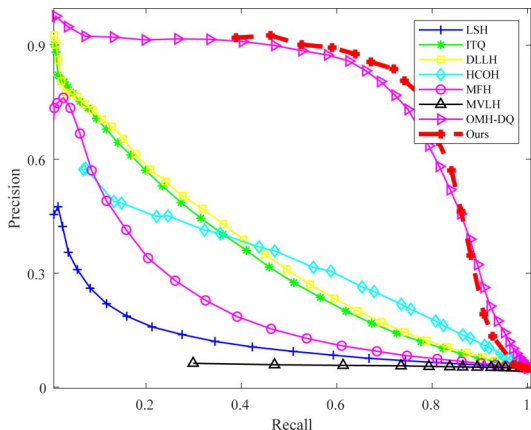
为了获得直观的视觉对比, 图 3 呈现一个多模态检索的例子. 几种对比方法和本文提出的 MFH-RH 返回的 5 个检索结果被展现在图 3 的右边. 检索结果是按从左到右进行排序的, 由于文本通常比较长, 因此检索结果中实例的文本数据被省去, 带有红色框的实例代表是正确相关的检索结果. 从图 3 中可以发现, ITQ 和 LSH 在前 5 个检索结果中未发现正确的检索结果; DLLE, HCOH, MFH, MVLH, OMH-DQ 在返回的结果中

排序为第四个或第五个实例才能找到相关正确的结果. 而本文提出的 MFH-RH 的相关结果的排序是第二个和第三个. 也就是说在前三个检索结果中, 本文提出的 MFH-RH 就能检索出与询问样本相关的检索结果. 图 3 中展现的检索例子表明, MFH-RH 方法优于其他的对比方法, 在多模态检索任务中能表现出优越的性能.

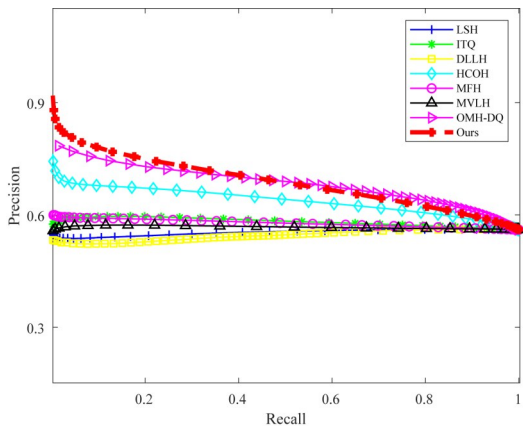
深度多模态哈希模型将异构多模态数据与当今



(a) Wiki



(b) Pascal Sentence



(c) MirFlickr

图2 所有方法在三个数据集上的PR曲线

先进的深度学习框架结合学习判别性的二值化编码特征,可以捕捉高层语义信息,在多模态哈希检索任务中取得了较高的检索精度.深度多视图哈希(Deep Multi-View Hashing, DMVH)^[33]和正交正则化多视图哈希(Deep Multi-modal Hashing with Orthogonal Regularization, DMHOR)^[34]都是无监督的深度方法.深度协作多视图哈希(Deep Collaborative Multi-

查询实例	方法	检索结果
 <p>Jordan returned to the NBA not as a player, but as part owner and President of Basketball Operations for the Washington Wizards.</p>	ITQ	
	LSH	
	DLLE	
	HCOH	
	MFH	
	MVLH	
	OMH-DQ	
	MFH-RH	

图3 Wiki数据集上多模态检索任务实例(由于篇幅限制,检索结果只呈现图片内容)

View Hashing, DCMVH)^[35]是一种有监督的深度多模态散列方法,它在类标签的监督下学习紧凑的散列码.语义驱动可解释的深度多模态哈希(Semantic-driven Interpretable Deep Multi-modal Hashing, SIDMH)^[36]在深度哈希结构中生成语义驱动的可解释哈希编码特征.在实验中,本文算法与上述深度哈希方法在Pascal Sentence数据集上进行实验对比,对比实验结果被呈现在表3中.与上述多模态深度模型相比,本文提出的算法取得了最好的检索性能.相比最佳的对比方法DCMVH,本文提出的MFH-RH在所设置的多种哈希长度的实验条件下获得平均0.91%的改善.实验结果表明,MFH-RH能够生成具有较强语义表达能力的哈希特征,进而有效地改善多模态哈希检索精度.

表3 不同方法在Pascal Sentence数据集上的性能比较

算法	Pascal Sentence			
	16 bit	32 bit	64 bit	128 bit
DMVH ^[33]	0.4832	0.5301	0.6010	0.6720
DMHOR ^[34]	0.5701	0.5930	0.6693	0.6860
DCMVH ^[35]	0.6793	0.7645	0.7734	0.7791
SIDMH ^[36]	0.6681	0.7479	0.7596	0.7660
MFH-RH	0.6867	0.7820	0.7820	0.7820

3.4 消融实验结果

本文基于Hadamard矩阵设计一种松弛的哈希中心去指导投影矩阵学习,以学习具有强大判别性的哈希函数.松弛的目的是增大不同类别数据间的边界.执行消融研究来验证松弛模型的有效性.式(4)是无松弛步骤的模型,记为MFH.在3个数据集上进行对比实

验,实验结果展现在表4中.从实验结果中看出MFH-RH在检索性能上优于MFH,尤其是在Wiki和Pascal Sentence数据集上.因此,这种基于Hadamard矩阵的松弛方法能够改善哈希检索的性能.对数据进行编码过程中,采用一种自适应的加权模式对多模态数据编码,以捕捉多媒体数据之间的差异性.为了验证其有效性,设计下面的对照实验.令“固定权值”表示在训练阶段学习获得的权值用于编码阶段.“动态权值”指的适应不同数据的动态权值.对比实验结果如图4所示.从实验结果中,“动态权值”模式相比“固定权值”性能得到了改善.相比MirFlickr数据集,在Wiki和Pascal Sentence上的提升更加明显.为了进一步研究其原因,通过实验观察数据编码过程中图片模态和文本模态对应权值变化趋势.从图4中可以看出Wiki和Pascal Sentence数据集上权值变化相比MirFlickr波动幅度更大.这说明Wiki和Pascal Sentence数据的差异性相对较大.结合图4、图5的实验结果,可以发现自适应的加权模式在数据差异较大的数据环境下更具鲁棒性.

表4 MFH-RH与非松弛的MFH方法的多媒体检索性能比较

数据集	方法	哈希编码长度			
		16 bit	32 bit	64 bit	128 bit
Wiki	MFH	0.6580	0.6674	0.6677	0.6752
	MFH-RH	0.6890	0.6927	0.6980	0.7005
Pascal Sentence	MFH	0.6519	0.7527	0.7527	0.7527
	MFH-RH	0.6867	0.7820	0.7820	0.7820
MirFlickr	MFH	0.6420	0.6867	0.6881	0.6868
	MFH-RH	0.6251	0.6900	0.6950	0.7063

3.5 参数敏感性分析

本文提出的模型存在2个惩罚参数 α 和 β .为了探究其对检索结果的影响,在实验过程中将它们值设置在一个较大的候选范围内 $\{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}, 1e^1\}$,来观察检索性能的变化.从图6中在3个数据集上的实验结果可以观察到,本文模型在一个较大区域内mAP值的变化较小.因此本文模型对超参数的敏感性不太大.

3.6 算法收敛性分析

在训练阶段,基于算法1的整个过程被重复直至算法收敛或者达到最大迭代次数.图7展示哈希长度设置为64 bit时,Wiki,Pascal Sentence和MirFlickr上的收敛曲线.在其他哈希长度设置下,曲线的变化趋势与64 bit长度的情况是相似的.从图7可以看出本文算法在10次迭代内就基本上趋于收敛.

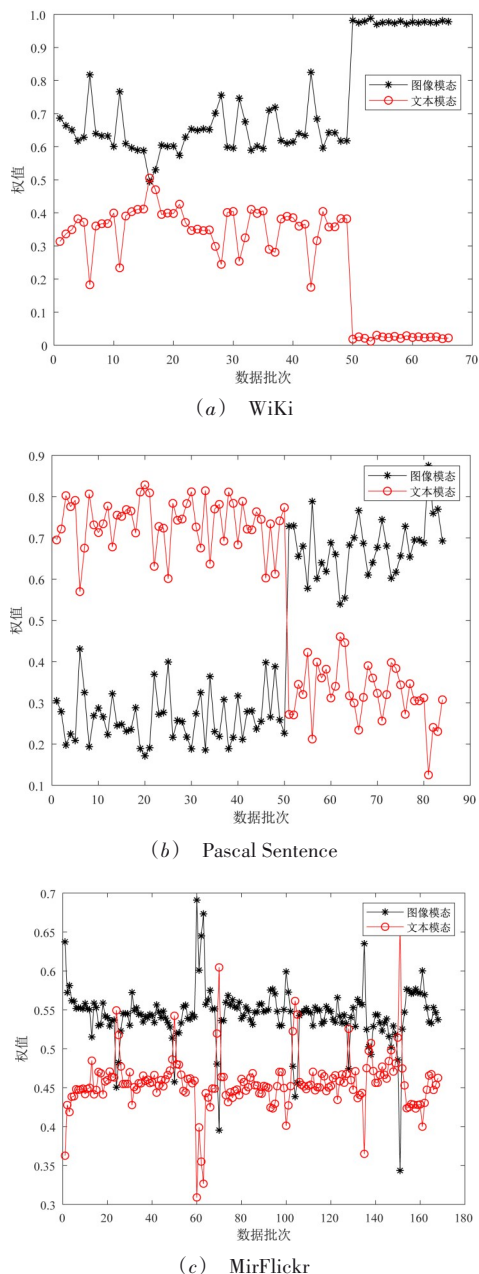


图4 数据编码过程中图片模态和文本模态对应权值变化趋势

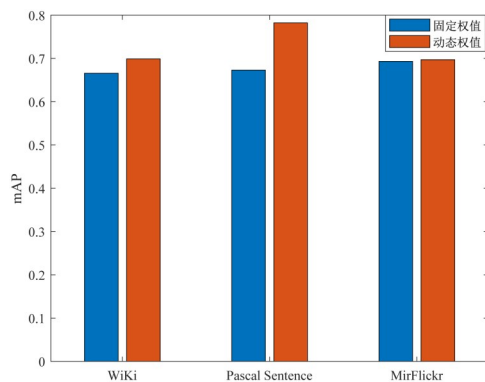
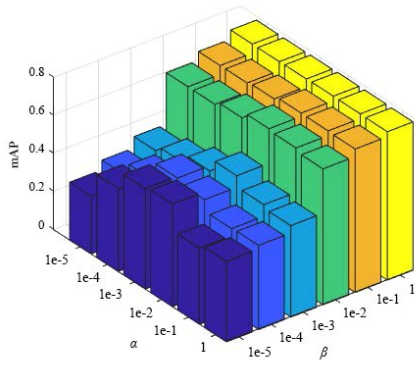
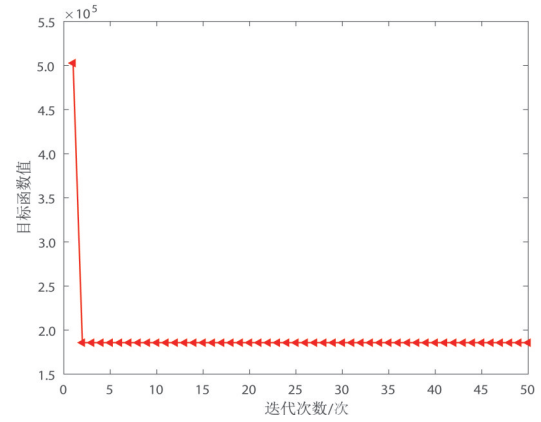


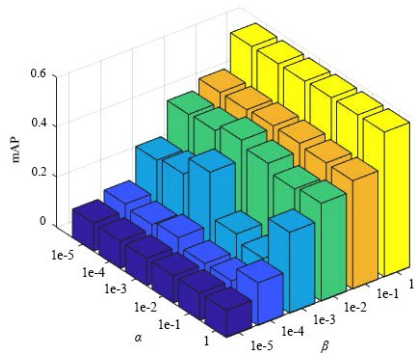
图5 自适应的动态权值和固定权值的对比实验结果



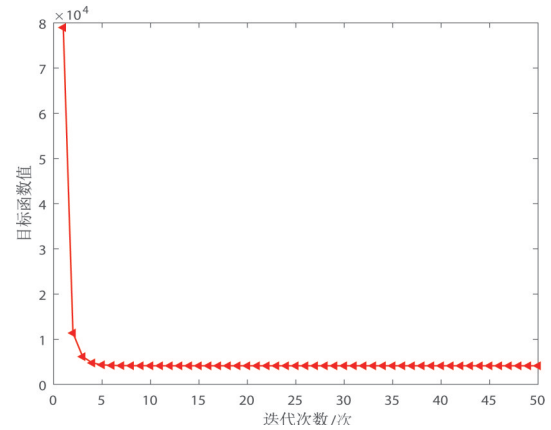
(a) Wiki



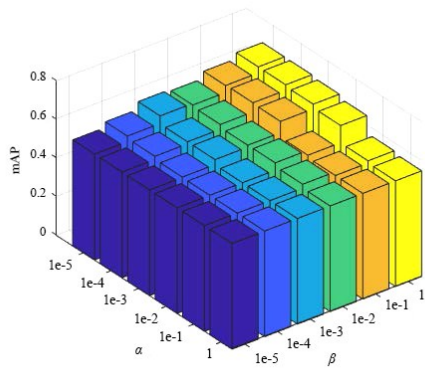
(a) Wiki



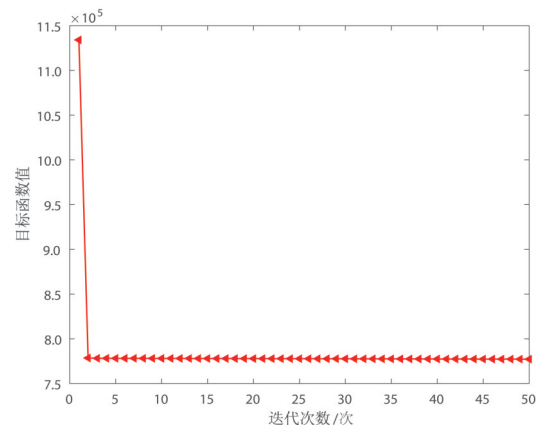
(b) Pascal Sentence



(b) Pascal Sentence



(c) MirFlickr



(c) MirFlickr

图6 不同 α 和 β 组合对应的mAP值

图7 训练阶段算法的收敛性曲线

4 结束语

本文提出了一种用于多媒体检索的新型多模态哈希学习方法. 该方法在类别信息的监督下利用Hadamard矩阵为数据生成目标哈希编码, 指导判别性的哈希函数学习. 在提出的框架中, 提出松弛严格的二值化目标编码为松弛变量矩阵, 这使得拥有更大的自由度去拟合目标矩阵, 从某种意义上增大了类间的间隔. 同

时保留类内的局部流形结构, 增强类内的紧凑性. 在编码阶段, 引入一种自适应的多模态加权方式, 有效地捕捉到数据的动态差异. 在3个公开的多模态数据集上进行对比实验, 在实验中本文提出的方法取得了很好的精度且超过了最新的方法. 实验结果表明, 基于松弛Hadamard矩阵的多模态哈希学习方法在多媒体检索中是非常有效的.

附录 1

松弛变量矩阵 \mathbf{E} 被定义为

$$\mathbf{E} = \begin{bmatrix} -1 - q_{11} & \cdots & +1 + q_{1r} \\ \vdots & \ddots & \vdots \\ +1 + q_{r1} & \cdots & -1 + q_{rr} \end{bmatrix}$$

$$\text{s.t. } q_{ij} \geq 0 \quad (22)$$

式(22)中, $q_{ij} \geq 0$ 有利于增大不同类别样本间的间距,

以学习判别性的投影矩阵. 其中 $\mathbf{H} = \begin{bmatrix} -1 & \cdots & +1 \\ \vdots & \ddots & \vdots \\ +1 & \cdots & -1 \end{bmatrix}$

通过一个辅助变量 \mathbf{Q} 和原始编码矩阵 \mathbf{H} 联合表示松弛变量矩阵 \mathbf{E} , 其中变量 \mathbf{Q} 被定义如下:

$$\mathbf{Q} = \begin{bmatrix} q_{11} & \cdots & q_{1r} \\ \vdots & \ddots & \vdots \\ q_{r1} & \cdots & q_{rr} \end{bmatrix}, \text{ s.t. } q_{ij} \geq 0 \quad (23)$$

因此 $\mathbf{E} = \mathbf{H} + \mathbf{H} \odot \mathbf{Q}$, 其中 \odot 是矩阵的 Hadamard 乘积.

参考文献

- [1] WANG J, SHEN H T, SONG J, et al. Hashing for similarity search: A survey[EB/OL]. (2014-08-13)[2022-02-12].<https://arxiv.org/abs/1408.2927>.
- [2] 李志欣, 凌锋, 张灿龙, 等. 融合两级相似度的跨媒体图像文本检索[J]. 电子学报, 2021, 49(2): 268-274.
LI Z X, LING F, ZHANG C L, et al. Cross-media image-text retrieval with two level similarity[J]. Acta Electronica Sinica, 2021, 49(2): 268-274. (in Chinese)
- [3] 李武军, 周志华. 大数据哈希学习: 现状与趋势[J]. 科学通报, 2015, 60(5): 485-490.
LI W J, ZHOU Z H. Learning to hash for big data: Current status and future trends[J]. Chinese Science Bulletin, 2015, 60(5): 485-490. (in Chinese)
- [4] 高文. “存得下, 查得快”拥抱多媒体大数据时代[J]. 创新科技, 2013, 25(6): 7.
GAO W. "Save it, check it quickly" Embrace the era of multimedia big data[J]. Innovation Science and Technology, 2013, 25(6): 7. (in Chinese)
- [5] 刘昊淼, 王瑞平, 山世光, 等. 基于离散优化的哈希编码学习方法[J]. 计算机学报, 2019, 42(5): 1149-1160.
LIU H M, WANG R P, SHAN S G, et al. Learning to hash with discrete optimization[J]. Chinese Journal of Computers, 2019, 42(5): 1149-1160. (in Chinese)
- [6] GONG Y, 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, 2012, 35(12): 2916-2929.
- [7] KULIS B, GRAUMAN K. Kernelized locality-sensitive hashing for scalable image search[C]//Proceedings of 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 2130-2137.
- [8] JI R, LIU H, CAO L, et al. Toward optimal manifold hashing via discrete locally linear embedding[J]. IEEE Transactions on Image Processing, 2017, 26(11): 5411-5420.
- [9] KOUTAKI G, SHIRAI K, AMBAI M. Hadamard coding for supervised discrete hashing[J]. IEEE Transactions on Image Processing, 2018, 27(11): 5378-5392.
- [10] JIN L, LI Z, PAN Y, et al. Weakly-supervised image hashing through masked visual-semantic graph-based reasoning[C]//Proceedings of the 28th ACM International Conference on Multimedia. Seattle: ACM, 2020: 916-924.
- [11] LI Z, TANG J, ZHANG L, et al. Weakly-supervised semantic guided hashing for social image retrieval[J]. International Journal of Computer Vision, 2020, 128: 2265-2278.
- [12] 姚涛, 孔祥维, 付海燕, 等. 基于映射字典学习的跨模态哈希检索[J]. 自动化学报, 2018, 44(8): 1475-1485.
YAO T, KONG X W, FU H Y, et al. Projective dictionary learning hashing for cross-modal retrieval[J]. Acta Automatica Sinica, 2018, 44(8): 1475-1485.
- [13] WANG D, GAO X, WANG X, et al. Label consistent matrix factorization hashing for large-scale cross-modal similarity search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(10): 2466-2479.
- [14] WANG D, WANG Q, GAO X. Robust and flexible discrete hashing for cross-modal similarity search[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(10): 2703-2715.
- [15] 刘昊鑫, 吴小俊, 庾骏. 联合哈希特征和分类器学习的跨模态检索算法[J]. 模式识别与人工智能, 2020, 33(2): 160-165.
LIU H X, WU X J, YU J. Joint hashing feature and classifier learning for cross-modal retrieval[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(2): 160-165. (in Chinese)
- [16] LIU H, JI R, WU Y, et al. Supervised matrix factorization

- for cross-modality hashing[C]//Proceedings of International Joint Conference on Artificial Intelligence. New York: IJCAI, 2016: 1767-1773.
- [17] LIN Z, DING G, HU M, et al. Semantics-preserving hashing for cross-view retrieval[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3864-3872.
- [18] 王锦荟, 金露, 李泽超, 等. 基于知识蒸馏的跨模态哈希[J/OL]. (2021-11-04)[2022-02-12]. <https://kns.cnki.net/kcms/detail/11.5844.TH.20220331.1444.008.html>.
WANG J Y, JIN L, LI Z C, et al. Cross-Modal Knowledge Distillation Hashing[J/OL]. (2021-11-04) [2022-02-12]. <https://kns.cnki.net/kcms/detail/11.5844.TH.20220331.1444.008.html>. (in Chinese)
- [19] LI Z, TANG J, MEI T. Deep collaborative embedding for social image understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(9): 2070-2083.
- [20] SONG J, YANG Y, HUANG Z, et al. Effective multiple feature hashing for large scale near duplicate video retrieval[J]. IEEE Transaction on Multimedia, 2013, 15(8): 1997-2008.
- [21] LIU X, HE J, LIU D, et al. Compact kernel hashing with multiple features[C]//Proceedings of the ACM International Conference on Multimedia. Seattle: ACM, 2012: 881-884.
- [22] SHEN X, SHEN F, SUN Q, et al. Multi-view latent hashing for efficient multimedia search[C]//Proceedings of the ACM International Conference on Multimedia, Seattle: ACM, 2015: 831-834.
- [23] SHEN X, SHEN F, LIU L, et al. Multiview discrete hashing for scalable multimedia search[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(5): 53-73.
- [24] LU X, ZHU L, CHENG Z, et al. Online multi-modal hashing with dynamic query-adaption[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris: ACM, 2019: 715-724.
- [25] SYLVESTER J J. LX. Thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1867, 34 (232): 461-475.
- [26] GIONIS A, INDYK P, MOTWANI R. Similarity search in high dimensions via hashing[C]// Proceedings of the International Conference on Very Large Date Bases. Sydney: IEEE, 1999: 518-529.
- [27] LIN M, JI R, LIU H, et al. Hadamard matrix guided online hashing[J]. International Journal of Computer Vision, 2020, 128(6): 2279-2306.
- [28] XIANG S, NIE F, MENG G, et al. Discriminative least squares regression for multiclass classification and feature selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(11): 1738-1754.
- [29] RASIWASIA N, COSTA P J, COVIELLO E, et al. A new approach to cross-modal multimedia retrieval[C]// Proceedings of the 18th ACM International Conference on Multimedia, Seattle: ACM, 2010: 251-260.
- [30] BLEI D M, NG A Y, JORDAN M I, et al. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(2): 993-1022.
- [31] RASHTCHIAN C, YOUNG P, HODOSH M, et al. Collecting image annotations using amazon's mechanical turk [C]//Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Los Angeles: ACM, 2010: 139-147.
- [32] WEI Y, ZHAO Y, LU C, et al. Cross-modal retrieval with CNN visual features: A new baseline[J]. IEEE Transactions on Cybernetics, 2016, 47(2): 449-460.
- [33] KANG Y, KIM S, CHOI S. Deep learning to hash with multiple representations[C]//Proceedings of the 12th IEEE Conference on Data Mining. Brussels: IEEE, 2012: 930-935.
- [34] WANG D, CUI P, OU M, et al. Deep multimodal hashing with orthogonal regularization[C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence. Buenos, Aires: IEEE, 2015: 2291-2297.
- [35] ZHU L, LU X, CHENG Z, et al. Deep collaborative multi-view hashing for large-scale image search[J]. IEEE Transactions on Image Processing, 2020, 29: 4643-4655.
- [36] LU X, LIU L, NIE L, et al. Semantic-driven Interpretable Deep Multi-modal Hashing for Large-scale Multimedia Re-

trieval[J]. IEEE Transactions on Multimedia, 2021, 23: 4541-4554.

作者简介



庾 骏 男,1990年生,湖南邵阳人. 工学博士. 现在郑州轻工业大学计算机与通信工程学院从事科研和教学工作. 研究领域为多媒体分析和检索、深度学习、模式识别.

E-mail: yujun@zzuli.edu.cn



黄 伟 男,1982年生,河南郑州人. 工学博士. 现为郑州轻工业大学计算机与通信工程学院副教授. 研究领域为模式识别、深度学习.

E-mail: hnhw235@163.com

张晓波 男,1987年生,河南郑州人. 工学博士. 现为郑州轻工业大学计算机与通信工程学院教师. 研究领域为稀疏优化、组合优化、数字信号处理.

E-mail: zxb@zzuli.edu.cn

尹贺峰 男,1989年生,江苏无锡人. 工学博士. 现为江南大学计算机与人工智能学院博士后. 研究领域为模式识别、机器学习.

E-mail: yin_hefeng@jiangnan.edu.cn