滕少华,黄文彪,张巍,等. 标签与样本双语义增强的跨模态检索[J]. 江西师范大学学报(自然科学版),2023,47(3):296-306.
TENG Shaohua, HUANG Wenbiao, ZHANG Wei, et al. The cross-modal hash with tag and sample semantic enhancements[J]. Journal of Jiangxi Normal University(Natural Science),2023,47(3):296-306.

文章编号:1000-5862(2023)03-0296-11

标签与样本双语义增强的跨模态检索

滕少华[1],黄文彪[1],张 巍[1],滕璐瑶[2]
(1. 广东工业大学计算机学院,广东 广州 510006;2. 广州番禺职业技术学院信息工程学院,广东 广州 511483)

摘要:针对目前大多数跨模态哈希检索方法无法捕获多标签信息和特征语义更深层的语义关系信息问题,该文提出了一种标签与样本双语义增强的跨模态检索框架. 首先,该框架将不同模态的高维数据映射到低维共享特征语义空间中,进行样本语义学习;其次,引入松弛变量到标签语义制约的哈希码学习函数中,通过最小化标签成对距离强化样本语义相似性哈希码学习,这样既保持了跨模态对应样本语义的关系,强化了哈希码的标签语义学习,又解决了实对称矩阵的求解及算法的收敛性问题;再次,进一步应用样本特征语义和标签语义增强哈希码的语义学习;最后,在 3 个常用的数据集上的实验结果表明该方法优于目前的方法.

关键词:标签与样本双语义增强;跨模态检索;标签语义
中图分类号:TP 311 **文献标志码**:A **DOI**:10.16357/j.cnki.issn1000-5862.2023.03.10

0 引言

近年来,多媒体数据开始爆炸式增长,且数据纬度高,存在于图片、文本、音频等不同模态中. 由于不同模态的数据可能具有相同的语义,所以出现了各种新的技术研究和应用需求,如稀疏表示学习[1]、跨域识别[2]和跨模态检索[3]等. 目前,跨模态检索受到广泛关注. 跨模态检索是通过一种模态样本来检索具有近似语义的另一种模态样本的检索技术,如文本到图像或图像到文本. 早年的跨模态检索方法是基于实值表示的,用于提升跨模态语义相关性,进而提高跨模态检索准确度. 然而,基于实值表示的跨模态检索方式的计算和存储成本较高,对大规模不同模态的多媒体数据执行高效和精确的跨模态最近邻搜索几乎是不可能的. 为了解决这个问题,研究者们提出了基于哈希的跨模态检索方法. 基于哈希的思想是将在高维空间中相似的样本映射到低维的汉明空间中对应的相似二进制哈希码表示,然后以通过简单的异或运算来计算汉明距离,根据汉明距离的大小来排序就可以实现近似搜索. 因此,计算和存储成本大大降低. 许多跨模态哈希方法[4-9]被相继提出.

现有的跨模态哈希方法主要可分为 2 类:无监督跨模态哈希和有监督跨模态哈希. 无监督跨模态哈希方法[5,7,10-13]旨在利用不同模态数据的潜在相关性来学习相似的哈希码. 与无监督跨模态哈希不同,有监督跨模态哈希方法[14-33]不仅利用原始数据的特征信息,还利用标签语义信息来学习相似的哈希码,这更有利于发现异构数据之间的相似关系. 因此,有监督跨模态方法比无监督跨模态方法获得了更高准确率的检索性能. 其中具有代表性的是有监督矩阵分解哈希(supervised matrix factorization hashing, SMFH)[14]、可拓展非对称的离散跨模态哈希(scalable asymmetric discrete cross-modal hashing, BATCH)[16]、语义相关最大化哈希(semantic correlation maximization hashing, SCMH)[19]、离散跨模态哈希(discriminative cross-modal hashing, DCH)[18]、子

收稿日期:2022-12-09
基金项目:国家自然科学基金(61972102)资助项目.
作者简介:滕少华(1962—),男,江西南昌人,教授,博士,博士生导师,主要从事大数据、数据挖掘、人工智能、模式识别、智能制造和网络安全方面的研究. E-mail:shteng@gdut.edu.cn

空间关系学习的跨模态哈希(subspace relation learning for cross-modal hashing, SRLCH)[25]等. 这些有监督跨模态哈希的方法以不同的方式将标签的语义信息嵌入哈希码中,如 SCMH[19]将标签语义集成到哈希码学习过程中来最大化数据相关性,但该方法使用松弛方案解决二进制约束,可能会导致较大的量化误差、次优哈希码和哈希函数. DCH[18]将标签语义信息学习当成分类问题解决,采用逐位优化的策略来学习哈希码;然而,逐位优化哈希码的方案会导致优化时间对哈希码的长度非常敏感,哈希码长度越长,优化时间越长. SMFH[14]直接将标签语义信息构造成一个 $n \times n$ 相似性矩阵,由此来指导学习相似的哈希码,但成对相似性矩阵对于哈希码的学习优化具有较大的时间成本和空间成本. SRLCH[25]将哈希码的学习当作标签信息的线性回归问题,利用标签信息来指导每一位哈希码的学习;但该方法只考虑标签的类别语义信息,忽略了原始数据之间的标签语义成对相似性. BATCH[16]将标签语义当作第 3 种模态来参与数据的矩阵分解过程,这可能会导致标签类别的区分性语义信息的丢失,从而学习到次优的哈希码. 综上所述,尽管有监督哈希方法取得了可喜的成就,但仍有一些问题需要进一步考虑:1)量化误差大. 为了解决离散哈希码优化问题,大多数跨模态哈希方法(如 SePH[6]、SMFH[14]、SCMH[19]、SRSH[20]、LCMFH[22])放松了对哈希码施加的离散约束. 2)大多数方法单一地考虑模态间的成对相似性,如 BATCH[16]通过分解成对相似性矩阵使得哈希码保持语义相似性信息,没有考虑标签语义的区分性信息. 3)大多数方法只考虑了标签的判别性语义. 如 SCRATCH[4]、SDDH[13]和 DCH[18]直接使用线性回归策略从 0 和 1 的标签学习哈希码,没有考虑标签语义的成对相似性信息.

为解决以上的问题,本文提出了一种标签与样本双语义增强的跨模态哈希检索方法. 该方法先将不同模态数据分解到低维共享潜在子空间中,再进行样本特征语义的学习;其次,在标签语义约束下,引入松弛变量,通过最小化标签成对距离增强样本语义相似的哈希码学习,保持了跨模态样本语义的相似性,也强化了哈希码的标签语义学习;最后,进一步应用样本特征语义和标签语义增强哈希码的语义学习. 这使得最终学习的哈希码在保持标签语义相似性的同时又具有判别性. 在 3 个广泛使用的数据集上的实验结果表明该方法优于目前的方法,特别是在多标签数据集上.

1　相关工作

如前文所述,相较于无监督哈希方法,有监督哈希方法能够结合语义标签来学习,因此更能够学习到保持语义相似性的哈希码. 如语义相关最大化哈希(SCMH)[19]将标签语义集成到哈希码学习过程中来最大化数据相关性,但该方法采用松弛策略解决哈希码的离散约束,带来了较大的量化误差;为了解决量化误差问题和避免使用成对相似性矩阵以解决时间复杂度高的问题,离散跨模态哈希(DCH)[18]保持离散约束并逐位迭代生成哈希码. 然而,这种逐位优化方式对哈希码长度较为敏感,也会导致比较大的优化时间复杂度问题. 为了解决这个问题,可拓展的离散矩阵分解哈希(SCRATCH)[4]使用矩阵分解和语义回归嵌入的策略来学习模态内和模态间的相似性,并离散地生成哈希码. 子空间关系学习的跨模态哈希(SRLCH)[25]将哈希码的学习当作标签信息的线性回归问题,利用数据点之间的相似性学习哈希码,同时利用标签信息来指导每一位哈希码的学习. 可拓展判别性离散跨模态哈希(SDDH)[13]直接通过矩阵分解来学习离散哈希码,利用标签信息回归到哈希码来提高哈希码的判别性. 但是这些方法都忽视了标签语义成对相似性的学习. 可拓展非对称的离散跨模态哈希(BATCH)[16]将成对相似度矩阵分解成 2 个标签矩阵的乘积,然后将标签信息嵌入哈希码的学习中,从而更好地利用了标签语义成对相似性和解决了离散优化问题. 但是该方法没有考虑标签语义的类别级别语义信息.

本文主要解决标签语义信息和特征语义信息的充分利用问题,以便捕获多标签信息和特征语义更深层的语义关系信息. 为此提出了一种标签与样本双语义增强的跨模态哈希方法,该方法不仅考虑标签的成对标签语义相似性,还考虑它的类别信息和原始数据的特征语义信息.

2　双语义增强的跨模态检索

本文中所使用的主要符号列于表 1. 算法的框架示意图如图 1 所示,主要分为哈希码的学习和哈希函数的学习. 在哈希码的学习中,该方法首先通过矩阵分解来学习不同模态数据之间的低维特征语义表示;然后使用分类矩阵 $\boldsymbol{Q}$ 来挖掘不同模态数据之间的标签语义的判别性信息;接着,通过线性

加权融合的方法来利用特征语义信息和标签判别性信息,以此来生成判别性哈希码矩阵 $\boldsymbol{B}$;最后,采用非对称策略,将特征语义信息作为一种辅助信息来指导哈希码对标签成对相似性信息的学习,这能够进一步保证语义相似性信息完全嵌入哈希码中. 在哈希函数学习中,基于已经学习好的哈希码,根据核化和线性回归来学习不同模态的哈希函数.

表 1 主要符号

符号	表示的意义
$\boldsymbol{X}^{\mathrm{T}}$	矩阵 $\boldsymbol{X}$ 的转置
$\boldsymbol{X}^{(t)}$	第 t 模态的特征矩阵
$\boldsymbol{X}^{(1)}$	文本的特征矩阵
$\boldsymbol{X}^{(2)}$	图像的特征矩阵
$\boldsymbol{B}$	哈希码矩阵
$\boldsymbol{L}$	标签矩阵
$\boldsymbol{G}$	2-范数归一化标签矩阵
$\boldsymbol{I}_n$	维度为 n 维、元素全为 1 的列向量
n	训练集实例的数量
r	哈希码的长度
c	标签的类别数
$\|\boldsymbol{X}\|_{\mathrm{F}}$	矩阵 $\boldsymbol{B}$ 的 Frobenius 范数
$\exp(x)$	x 的指数函数 e^x
$\mathrm{tr}(\boldsymbol{X})$	矩阵 $\boldsymbol{X}$ 的迹
$\mathrm{sgn}(x)$	x 的符号函数
$\phi(x)$	x 的核化

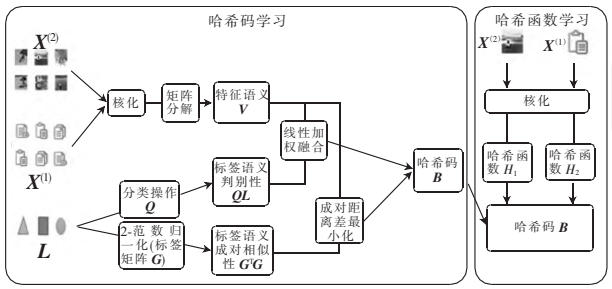

图 1 算法框架示意图

2.1 哈希码学习

2.1.1 *样本特征语义信息学习* 矩阵分解[5]是在学习原始数据中潜在特征的最有用工具之一. 许多跨模态方法通过这一技术直接从不同模态中学习共同的潜在特征,没有考虑数据的非线性特征,从而导致学习的潜在特征的非线性不足. 为了解决这个问题,本文先对原始数据进行核化,进一步提取原始数据的非线性特征,从而更好地提取有用的潜在特征信息. 此外,为了减少数据特征之间的冗余和使得特征学习更稳定收敛,在矩阵分解过程中引入了正交约束和平衡约束:

$$\min \sum_{t=1}^{2} \alpha \|\phi(\boldsymbol{X}^{(t)}) - \boldsymbol{U}_t\boldsymbol{V}\|_{\mathrm{F}}^2 + \delta\|\boldsymbol{U}_t\|_{\mathrm{F}}^2, \quad (1)$$

$$\text{s. t. } \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = n\boldsymbol{I}_r, \boldsymbol{V}_{1n} = \boldsymbol{0}_r$$

其中 α 和 δ 是超参数,$\boldsymbol{U}_t \in \mathbf{R}^{k_t \times r}$ 是基向量,$\boldsymbol{V} \in \mathbf{R}^{r\times n}$ 是潜在特征表示,$\phi(\cdot)$ 是径向基函数(RBF)核,即 $\phi(x) = (\exp(\|x - \boldsymbol{a}_1^{(t)}\|^2/(-2\sigma_t^2)), \cdots, \exp(\|x - \boldsymbol{a}_{k_t}^{(t)}\|^2/(-2\sigma_t^2)))$,其中 $\boldsymbol{a}_i^{(t)} \in \mathbf{R}^{d_t}(i = 1,2,\cdots,k_t)$ 是 k_t 训练样本中的锚点数目,σ_t 是第 t 模态的核宽度.

2.1.2 *标签语义制约的哈希码学习* 许多有监督哈希方法会构造一个 $n \times n$ 大小的语义相似性矩阵来利用标签信息,从而确保语义相似的原始数据能学习到相似的哈希码. 然而,$n \times n$ 大小的成对相似性矩阵会带来较低的优化效率和较大的存储成本. 为了攻克这个难题,BATCH[16]提出了一种通过最小化标签成对距离和哈希码成对距离之间的距离差的学习方案:

$$\min \|\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} - r\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\|_{\mathrm{F}}^2, \quad (2)$$

$$\text{s. t. } \boldsymbol{B} \in \{-1,1\}^{r\times n},$$

其中 $\boldsymbol{G}$ 为 2-范数归一化后的标签矩阵,其元素为 $\boldsymbol{G}_i = \boldsymbol{L}_i/\|\boldsymbol{L}_i\|$.

通过式(2),虽然可以保证以比较高的效率来学习优化高质量的哈希码,但是在式(2)中对称的离散哈希码的优化是一个 NP 难题. 为了解决这个优化难题,许多跨模态哈希方法采用松弛的策略,将二进制取值变成连续实值取值参与算法的整个优化过程,待优化后再将连续取值通过符号函数映射为二进制取值. 然而,这将导致较大的量化误差. 为了解决这个问题,本文引入一个非对称策略,利用特征语义信息 $\boldsymbol{V}$ 替代其中一个 $\boldsymbol{B}$,得到

$$\min \|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{B} - r\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\|_{\mathrm{F}}^2, \quad (3)$$

$$\text{s. t. } \boldsymbol{B} \in \{-1,1\}^{r\times n}, \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = n\boldsymbol{I}_r, \boldsymbol{V}_{1n} = \boldsymbol{0}_r.$$

通过这种非对称策略,将量化的操作局限在算法优化的每一次迭代中,通过 $\boldsymbol{V}$ 不断迭代的更新来对 $\boldsymbol{B}$ 进行修改,使得 $\boldsymbol{B}$ 达到松弛并稳定更新的作用. 这不仅可以缓解采用松弛策略所能带来的量化误差问题,而且可以将特征语义信息作为一个监督信息来辅助标签成对相似性信息嵌入所学习的哈希码中.

2.1.3 *双语义增强的哈希码学习* 本文在式(3)中采用非对称策略,利用特征语义信息 $\boldsymbol{V}$ 去替换其中一个 $\boldsymbol{B}$,使得优化问题变得快速可解. 但是这存在

以下问题:1)$\boldsymbol{V}$ 和 $\boldsymbol{B}$ 之间存在一定的量化误差;2)由于标签语义信息是通过余弦相似度计算成一种成对相似性信息来监督哈希码的学习,使得标签语义之间的区分性信息得不到有效的利用. 为了解决这 2 个问题,本文提出了一种通过线性加权融合策略,将特征语义信息和标签语义判别性信息融合起来,进一步增强哈希码在语义上的学习,可表示为

$$\min \lambda \| \boldsymbol{B}-\eta \boldsymbol{V}-(1-\eta)\boldsymbol{Q}\boldsymbol{L} \|_{\mathrm{F}}^{2}, \tag{4}$$

其中 η 和 λ 是超参数,$\boldsymbol{Q}$ 矩阵是一个分类器,用于将标签语义信息回归到哈希码,从而使得标签语义的判别性信息能够被保持在所学习的哈希码中.

2.1.4　*整体目标函数*　算法的整体目标函数是将式(1)、式(3)和式(4)联合起来优化,得

$$\begin{aligned}&\min \sum_{t=1}^{2} \alpha \| \phi(\boldsymbol{X}^{(t)})-\boldsymbol{U}_t\boldsymbol{V} \|_{\mathrm{F}}^{2}+\| \boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}-r\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} \|_{\mathrm{F}}^{2}+\\ &\lambda \| \boldsymbol{B}-\eta \boldsymbol{V}-(1-\eta)\boldsymbol{Q}\boldsymbol{L} \|_{\mathrm{F}}^{2}+\delta \| \boldsymbol{U}_t \|_{\mathrm{F}}^{2}, \qquad (5)\\ &\text{s.t. } \boldsymbol{B}\in\{-1,1\}^{r\times n}, \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r.\end{aligned}$$

2.1.5　*优化算法*　直接同时优化具有 5 个变量($\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{V},\boldsymbol{B},\boldsymbol{Q}$)的非凸目标函数是非常困难的. 然而可以每次只优化一个变量,然后固定其他变量不变,通过这种迭代更新来为每一个变量获得一个近似解. 具体的迭代过程如下:

1)固定 $\boldsymbol{U}_2$、$\boldsymbol{V}$、$\boldsymbol{B}$、$\boldsymbol{Q}$,更新 $\boldsymbol{U}_1$,则式(5)可写为

$$\min \alpha \| \phi(\boldsymbol{X}^{(1)})-\boldsymbol{U}_1\boldsymbol{V} \|_{\mathrm{F}}^{2}+\delta \| \boldsymbol{U}_1 \|_{\mathrm{F}}^{2}, \tag{6}$$

展开式(6)并对 $\boldsymbol{U}_1$ 求导,令其导数为 0,在 $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r$ 约束下可得到闭式解

$$\boldsymbol{U}_1=\alpha\phi(\boldsymbol{X})^{(1)})\boldsymbol{V}^{\mathrm{T}}(\alpha n\boldsymbol{I}_r+\delta\boldsymbol{I}_r)^{-1}. \tag{7}$$

2)固定 $\boldsymbol{U}_1$、$\boldsymbol{V}$、$\boldsymbol{B}$、$\boldsymbol{Q}$,更新 $\boldsymbol{U}_2$,则式(5)可写为

$$\min \alpha \| \phi(\boldsymbol{X}^{(2)})-\boldsymbol{U}_2\boldsymbol{V} \|_{\mathrm{F}}^{2}+\delta \| \boldsymbol{U}_2 \|_{\mathrm{F}}^{2}, \tag{8}$$

展开式(8)并对 $\boldsymbol{U}_1$ 求导,令其导数为 0,在 $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r$ 约束下可得闭式解

$$\boldsymbol{U}_2=\alpha\phi(\boldsymbol{X}^{(2)})\boldsymbol{V}^{\mathrm{T}}(\alpha n\boldsymbol{I}_r+\delta\boldsymbol{I}_r)^{-1}. \tag{9}$$

3)固定 $\boldsymbol{U}_1$、$\boldsymbol{U}_2$、$\boldsymbol{B}$、$\boldsymbol{Q}$,更新 $\boldsymbol{V}$,则式(5)可写为

$$\begin{aligned}&\min \sum_{t=1}^{2} \alpha \| \phi(\boldsymbol{X}^{(t)})-\boldsymbol{U}_t\boldsymbol{V} \|_{\mathrm{F}}^{2}+\| \boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}-r\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} \|_{\mathrm{F}}^{2}+\\ &\lambda \| \boldsymbol{B}-\eta \boldsymbol{V}-(1-\eta)\boldsymbol{Q}\boldsymbol{L} \|_{\mathrm{F}}^{2},\\ &\text{s.t. } \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r. \qquad (10)\end{aligned}$$

为了解决上述问题,本文将式(10)中的目标函数在 $\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r$ 和 $\boldsymbol{B}\in\{-1,1\}^{r\times n}$ 约束下转为矩阵的迹的形式,最终简化为

$$\begin{aligned}&\max \mathrm{tr}((\lambda\eta\boldsymbol{B}-(1-\eta)\boldsymbol{Q}\boldsymbol{L}+r\boldsymbol{B}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\alpha\boldsymbol{U}_1{}^{\mathrm{T}}\cdot\\ &\phi(\boldsymbol{X}^{(1)})+\alpha\boldsymbol{U}_2{}^{\mathrm{T}}\phi(\boldsymbol{X}^{(2)}))\boldsymbol{V}^{\mathrm{T}}),\\ &\text{s.t. } \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}=n\boldsymbol{I}_r, \boldsymbol{V}_{1n}=\boldsymbol{0}_r.\end{aligned}$$

本文定义 $\boldsymbol{P}=\boldsymbol{I}_n-\boldsymbol{I}_n\boldsymbol{I}_n{}^{\mathrm{T}}/n$,$\boldsymbol{Z}=\lambda\eta\boldsymbol{B}-(1-\eta)\cdot\boldsymbol{Q}\boldsymbol{L}+r\boldsymbol{B}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\alpha\boldsymbol{U}_1{}^{\mathrm{T}}\phi(\boldsymbol{X}^{(1)})+\alpha\boldsymbol{U}_2{}^{\mathrm{T}}\phi(\boldsymbol{X}^{(2)})$.

然后,$\boldsymbol{Z}\boldsymbol{P}\boldsymbol{Z}^{\mathrm{T}}$ 用于奇异值分解,具体式子为

$$\boldsymbol{Z}\boldsymbol{P}\boldsymbol{Z}^{\mathrm{T}}=(\boldsymbol{Q}\quad\ddot{\boldsymbol{Q}})\begin{pmatrix}\boldsymbol{\Omega}&0\\0&0\end{pmatrix}(\boldsymbol{Q}\quad\ddot{\boldsymbol{Q}})^{\mathrm{T}},$$

其中 $\boldsymbol{\Omega}\in\mathbf{R}^{r'\times r'}$ 是由正特征值构成的对角矩阵,$\boldsymbol{Q}\in\mathbf{R}^{r'\times r'}$ 和 $\ddot{\boldsymbol{Q}}\in\mathbf{R}^{(r-r')\times r'}$ 分别对应于正特征值的特征向量构成的矩阵和零特征值的 $r-r'$ 个特征向量构成的矩阵,r' 是 $\boldsymbol{Z}\boldsymbol{P}\boldsymbol{Z}^{\mathrm{T}}$ 的秩. 对 $\ddot{\boldsymbol{Q}}$ 进行 Gram-Schmidt 正交化处理来得到 $\bar{\boldsymbol{Q}}\in\mathbf{R}^{r\times(r-r')}$. 进一步定义 $\boldsymbol{J}=\boldsymbol{P}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\Omega}^{-1}/2$ 和一个随机正交矩阵 $\bar{\boldsymbol{J}}\in\mathbf{R}^{n\times(r-r')}$. 最后根据文献[34],可得 $\boldsymbol{V}$ 的最优解为

$$\boldsymbol{V}=\sqrt{n}(\boldsymbol{Q}\quad\bar{\boldsymbol{Q}})(\boldsymbol{J}\quad\bar{\boldsymbol{J}})^{\mathrm{T}}. \tag{12}$$

4)固定 $\boldsymbol{U}_1$、$\boldsymbol{U}_2$、$\boldsymbol{V}$、$\boldsymbol{B}$,更新 $\boldsymbol{Q}$,则式(5)可写为

$$\min \lambda \| \boldsymbol{B}-\eta \boldsymbol{V}-(1-\eta)\boldsymbol{Q}\boldsymbol{L} \|_{\mathrm{F}}^{2}. \tag{13}$$

展开式(13)并对 $\boldsymbol{Q}$ 求导,令其导数为 0,可得到闭式解

$$\boldsymbol{Q}=(\lambda(1-\eta)\boldsymbol{B}-\lambda\eta(1-\eta)\boldsymbol{V})(\lambda(1-\eta)^{2}\boldsymbol{L})^{-1}. \tag{14}$$

5)固定 $\boldsymbol{U}_1$、$\boldsymbol{U}_2$、$\boldsymbol{V}$、$\boldsymbol{Q}$,更新 $\boldsymbol{B}$,则式(6)可写为

$$\min \| \boldsymbol{V}^{\mathrm{T}}\boldsymbol{B}-r\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} \|_{\mathrm{F}}^{2}+\lambda \| \boldsymbol{B}-\eta \boldsymbol{V}-(1-\eta)\cdot\boldsymbol{Q}\boldsymbol{L} \|_{\mathrm{F}}^{2}, \tag{15}$$

展开式(15),可以等价地转化为

$$\begin{aligned}&\max \mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}(r\boldsymbol{V}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}-\lambda\eta\boldsymbol{V}-\lambda(1-\eta)\boldsymbol{Q}\boldsymbol{L}), \qquad (16)\\ &\text{s.t. } \boldsymbol{B}\in\{-1,1\}^{r\times n}.\end{aligned}$$

因此,式(16)的最优解为

$$\boldsymbol{B}=\mathrm{sgn}(r\boldsymbol{V}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}-\lambda\eta\boldsymbol{V}-\lambda(1-\eta)\boldsymbol{Q}\boldsymbol{L}). \tag{17}$$

2.2　哈希函数学习

如前所述,本文提出的算法为 2 步哈希方法. 在哈希码的学习阶段,所学习得到的哈希码矩阵 $\boldsymbol{B}$ 用于指导对应模态的哈希函数的学习. 在学习哈希函数之前,本文先对原始数据进行核化处理以便捕捉数据间的非线性特征. 第 t 模态的目标哈希函数为

$$\min \| \boldsymbol{B}-\boldsymbol{W}_t\phi(\boldsymbol{X}^{(t)}) \|_{\mathrm{F}}^{2}+\theta \| \boldsymbol{W}_t \|_{\mathrm{F}}^{2},$$

其中 θ 是超参数,且 $\boldsymbol{W}_t$ 的最优解为

$$\boldsymbol{W}_t=\boldsymbol{B}\phi(\boldsymbol{X}^{(t)})^{\mathrm{T}}(\phi(\boldsymbol{X}^{(t)})^{\mathrm{T}}+\theta\boldsymbol{I}_{k_t})^{-1}. \tag{18}$$

因此,给定一个第 t 模态的查询样本(如 $\boldsymbol{q}^{(t)}$),本文可以通过以下哈希函数得到对应的哈希码:

$$F_t(\boldsymbol{q}^{(t)})=\mathrm{sgn}(\boldsymbol{W}_t\phi(\boldsymbol{q}^{(t)})).$$

2.3　跨模态检索算法

由 2.1 小节和 2.2 小节的分析,可以形成本文模型的算法.

算法 1　标签与样本双语义增强的跨模态检索

算法.

输入:训练集 $\phi(\boldsymbol{X}^{(t)})(t=\{1,2\})$,标签值 $\boldsymbol{L}$,2-范数归一化标签矩阵 $\boldsymbol{G}$,哈希码长度 r,迭代次数 T,参数 α、λ、γ、δ、θ.

输出:哈希码 $\boldsymbol{B}$, 哈希函数 $F_t(\cdot)$.

/*第 1 步:哈希码学习阶段*/

1)初始化:随机初始化 $\boldsymbol{U}_1$,$\boldsymbol{U}_2$,$\boldsymbol{V}$,$\boldsymbol{B}$,$\boldsymbol{Q}$,

重复:

2)通过式(7)更新 $\boldsymbol{U}_1$,

3)通过式(9)更新 $\boldsymbol{U}_2$,

4)通过式(12)更新 $\boldsymbol{V}$,

5)通过式(14)更新 $\boldsymbol{Q}$,

6)通过式(17)更新 $\boldsymbol{B}$,

直到达到收敛或者最大迭代次数 T.

/*第 2 步:哈希函数学习阶段*/

7)通过式(18)更新学习哈希函数 $F_t(\cdot)$,返回哈希码 $\boldsymbol{B}$,哈希函数 $F_t(\cdot)$.

2.4 算法复杂性分析

式(7)和式(9)的时间复杂度为 $O(T(dnr+r^3+dr^2))$. 式(12)的时间复杂度为 $O(T(dr^2+nr^2+r^3+rdn+rn^2+rn+rcn))$. 式(14)的时间复杂度为 $O(T(nr+cn^2+rcn))$. 式(17)的时间复杂度为 $O(T(n^2r++rn+crn))$. r 为哈希码长度,n 为训练数据个数,c 为类别总数,T 为算法迭代次数,d 表示使用 $\phi(\cdot)$后的特征维数.

3 实验与结果分析

为了评估本文算法的性能和有效性,在 3 个基准数据集上将本文算法与 7 种近期跨模态哈希方法进行了充分的对比实验.

3.1 数据集

LabelMe[35]:它包含 2 688 个图像文本样本对. 数据集有 8 个类别,每个样本对都属于其中一个类别. 文本由 245 维词频特征向量表示,图像由 512 维 Gist 特征向量表示. 在该数据集中,随机选择 2 016 个样本对作为训练集,其余 672 个样本对用作测试集.

MIRFlickr[36]:它包含 25 000 个从 Flickr 网站下载的图像-文本样本对. 每个样本对用 24 个标记中的 1 个或多个标记. 图像由 PCA 降维后的 150 维边缘直方图特征向量表示,相应的文本也由 PCA 降维度后的 500 维特征向量表示. 随机选择 3% 的样本对作为查询集,65% 的样本对用作训练集.

NUS-WIDE[37]:它由 Flickr 获取的 269 648 个图像-文本样本对组成. 每个样本对至少有 1 个标签. 所有样本对可分为 81 类. 在这个实验中,选择了 10 个最频繁的标签,共有 186 577 个样本对. 在每种情况下,图像由 500 维视觉袋 SIFT 直方图特征向量表示,相应的文本由 1 000 维索引特征向量表示. 在 NUS-WIDE 中随机选择 20 000 个训练样本对和 2 000个测试样本对来训练和测试所有方法.

3.2 对比方法及其简介

为了评估本文算法的性能,将本文算法与 7 种近期跨模态哈希方法进行了比较,包括 2 种无监督方法(CMFH[5] 和 JIMFH[15])以及 5 种有监督方法(DCH[18]、SCRATCH[4]、SRLCH[25]、BATCH[16] 和 SDDH[13]).

CMFH[5] 利用矩阵分解技术对原始数据学习潜在特征,并由此学习对应模态哈希码.

JIMFH[16] 利用矩阵分解技术对原始数据学习对应模态的潜在特征和所有模态共享的潜在特征,从而提高所学习的哈希码的质量.

DCH[18] 通过直接学习哈希码上的线性映射来预测类信息,而类信息在本质上被视为线性分类问题,并且可以获得有区分性的哈希码.

SCRATCH[4] 使用矩阵分解和语义回归嵌入策略,以保持模态间的相似性,并以离散的方式来生成哈希码.

SRLCH[25] 直接将标签信息视为高级特征,学习从标签空间到汉明空间的线性变换,并将标签信息回归到哈希码.

BATCH[16] 通过最小化标签成对距离和哈希码成对距离之间的距离差来提高哈希码的标签语义相似性保持能力.

SDDH[13] 直接通过矩阵分解来学习哈希码,并且将标签信息回归到哈希码,进一步提高哈希码的判别能力.

所有对比方法都是采用公开的源代码. 所有对比方法的参数都是根据笔者的建议设置的. 在实验中,设置了 2 个跨模态检索任务:图像检索文本和文本检索图像. 所有实验都在具有 Intel(R)Core(TM) i9-9900K CPU@3.60 GHz、32 GB RAM 的工作站上进行. 本文算法在 3 个数据集上的超参数设置如下:$\alpha=1\times10^{-5}$,$\lambda=1\times10^{-2}$,$\eta=0.5$,$\delta=1\times10^{-2}$,$\theta=0.1$. 哈希码长度 r 分别设置为 32、64、96 和 128. 最大迭代次数设置为 5.

3.3　评价指标

为了评估每种方法的性能,使用了 2 种广泛使用的评估度量:平均精度均值(M_{AP})和 precision-recall 曲线. 平均精度均值的计算需要先求平均精度的值,具体来说,有查询 $\boldsymbol{q}$ 和检索实例列表 $\boldsymbol{R}$,$\boldsymbol{q}$ 的平均精度 A_P 定义如下:

$$A_P = \frac{1}{N}\sum_{r=1}^{n} P(r)\theta(r) \ ,$$

其中 N 是在检索数据库中与 $\boldsymbol{q}$ 相关的实例数量,n 是实例列表 $\boldsymbol{R}$ 的实例数量,$P(r)$ 表示被检索到的前 r 个实例的精度. $\theta(r)=1$ 表示检索到的第 r 个实例与查询 $\boldsymbol{q}$ 相关,否则 $\theta(r)=0$. 最后,计算所有查询实例 A_P 的值,再对其取平均以便获得 M_{AP}.

对于评估指标 M_{AP},其值越大代表着算法性能越好. 为了评估计算时间成本,本文还记录了具有不同代码长度的所有方法的训练时间.

3.4　实验结果分析

表 2、表 3 和表 4 总结了本文算法和对比方法在 3 个数据集上的 M_{AP} 值. 从表 2、表 3 和表 4 可以得出如下结论:

1)本文算法的表现在文本检索图像任务中要优于所有对比方法,特别是在多标签数据集上,精度提升了 3.00%. 在 LabelMe 数据集图像检索文本和文本检索图像任务中,本文算法在 $r=32$ 时略弱于 SDDH 和 BATCH,这表明了本文算法是有效的.

表 2　各方法在 LabelMe 数据集上的 M_{AP} 值

任务	方法	r			
		32	64	96	128
图像检索文本	CMFH	0.474 3	0.478 5	0.484 6	0.488 6
	JIMFH	0.590 2	0.597 8	0.587 4	0.575 4
	DCH	0.812 0	0.832 8	0.844 7	0.834 2
	SCRATCH	0.876 2	0.866 8	0.863 3	0.854 1
	SRLCH	0.870 5	0.885 5	0.883 9	0.889 6
	BATCH	0.886 1	0.889 6	0.890 9	0.891 2
	SDDH	0.896 7	0.884 5	0.891 4	0.891 1
	本文算法	0.894 9	0.893 5	0.898 7	0.892 1
文本检索图像	CMFH	0.525 3	0.518 5	0.527 9	0.526 1
	JIMFH	0.690 8	0.710 1	0.610 0	0.685 6
	DCH	0.911 9	0.913 7	0.919 0	0.919 8
	SCRATCH	0.922 0	0.913 7	0.910 3	0.909 3
	SRLCH	0.920 2	0.920 3	0.925 0	0.924 7
	BATCH	0.931 8	0.926 8	0.927 4	0.928 2
	SDDH	0.928 4	0.924 7	0.931 0	0.929 3
	本文算法	0.927 9	0.933 1	0.931 1	0.936 6

注:32、64、96、128 为哈希码长度 r 的取值. 下文同.

2)有监督哈希方法比无监督 CMFH 和 JIMFH 方法获得更好的性能,这验证了语义标签对于高质量和高精确的哈希码的学习具有较大的重要性.

3)在 LabelMe 数据集 32 位图像检索文本和文本检索图像任务中,本文算法的 M_{AP} 比最优方法的 M_{AP} 差距在 0.38% 内,而在 128 位下分别提升至 0.90%、0.84%. 在 MIRFlickr 数据集 32 位同任务中,M_{AP} 分别提升 2.82%、0.68%,在 128 位下分别提升 3.00%、0.85%. 在 NUS-WIDE 数据集 32 位同任务下,M_{AP} 分别提升 3.27%、0.56%,在 128 位下分别提升 4.57%、1.72%. 可以看出本文算法在多标签数据集上的表现更佳,而在单标签数据集中的表现与最优对比算法相似. 这可能的原因是:在单标签数据集上,图像特征信息和文本特征信息很可能在标签语义上比较接近,语义关系结构没那么复杂;而相对于单标签数据集,多标签数据集蕴含更多的语义信息,语义关系结构更复杂. 本文算法可以较好地通过双语义增强的策略来挖掘多标签数据更深度的语义信息,更能够学习到精确的哈希码,从而实现了更好的检索性能.

表 3　各方法在 MIRFlickr 数据集上的 M_{AP} 值

任务	方法	r			
		32	64	96	128
图像检索文本	CMFH	0.585 4	0.584 5	0.586 2	0.585 4
	JIMFH	0.652 0	0.659 0	0.671 1	0.669 0
	DCH	0.684 5	0.681 0	0.690 8	0.691 1
	SCRATCH	0.705 7	0.723 8	0.728 3	0.726 4
	SRLCH	0.591 2	0.659 0	0.635 3	0.631 7
	BATCH	0.745 8	0.749 0	0.749 9	0.752 3
	SDDH	0.729 5	0.741 8	0.746 2	0.748 9
	本文算法	0.774 0	0.774 8	0.789 0	0.782 3
文本检索图像	CMFH	0.596 9	0.594 6	0.596 5	0.593 9
	JIMFH	0.692 8	0.697 1	0.704 8	0.702 7
	DCH	0.763 6	0.773 3	0.783 0	0.798 5
	SCRATCH	0.783 0	0.804 1	0.809 7	0.808 3
	SRLCH	0.634 7	0.712 9	0.687 3	0.682 3
	BATCH	0.834 2	0.839 5	0.842 9	0.844 4
	SDDH	0.801 8	0.812 8	0.828 0	0.833 2
	本文算法	0.841 0	0.847 9	0.855 5	0.852 9

表 4 各方法在 NUS-WIDE 数据集上的 M_{AP} 值

任务	方法	r			
		32	64	96	128
图像检索文本	CMFH	0.387 7	0.386 2	0.386 2	0.385 7
	JIMFH	0.488 1	0.497 1	0.505 7	0.505 2
	DCH	0.588 9	0.587 6	0.607 6	0.605 4
	SCRATCH	0.640 7	0.647 9	0.648 7	0.652 0
	SRLCH	0.602 1	0.610 7	0.627 2	0.624 8
	BATCH	0.642 8	0.656 1	0.659 3	0.660 6
	SDDH	0.639 6	0.664 3	0.656 6	0.661 6
	本文算法	0.675 5	0.691 9	0.686 0	0.707 3
文本检索图像	CMFH	0.398 1	0.403 9	0.405 8	0.405 5
	JIMFH	0.524 2	0.533 3	0.542 5	0.550 0
	DCH	0.709 4	0.700 2	0.734 7	0.730 2
	SCRATCH	0.755 2	0.761 4	0.765 9	0.775 5
	SRLCH	0.712 3	0.727 7	0.741 2	0.741 2
	BATCH	0.784 4	0.797 8	0.795 0	0.800 0
	SDDH	0.774 9	0.794 3	0.793 4	0.793 6
	本文算法	0.790 0	0.807 1	0.801 1	0.817 2

4)对于大多数方法,检索精度随着哈希码长度的增加而提升. 这可能的原因是:低位哈希码不足以编码丰富的语义信息,从而导致学习到的哈希码质量不高;相反地,更长的哈希码就可以编码更多的语义信息.

图 2 ~ 图 7 展示了本文算法和对比方法的 precision-recall(精确率-召回率)曲线. 在理论上,M_{AP} 值越高,对应的 precision-recall 曲线也越高. 从图 2 ~ 图 7 可以看出,本文算法基本上高于所有对比方法,这与表 2、表 3 和表 4 的实验结果是一致的.

由以上实验结果分析可知,本文算法比目前先进的方法更有良好的检索性能. 这说明本文算法是可以学习到更加精确的哈希码.

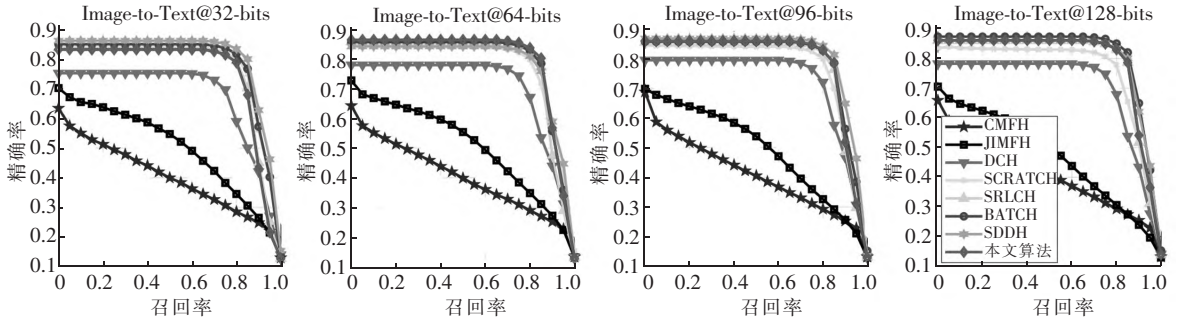

图 2 LabelMe 图像检索文本 PR 曲线图

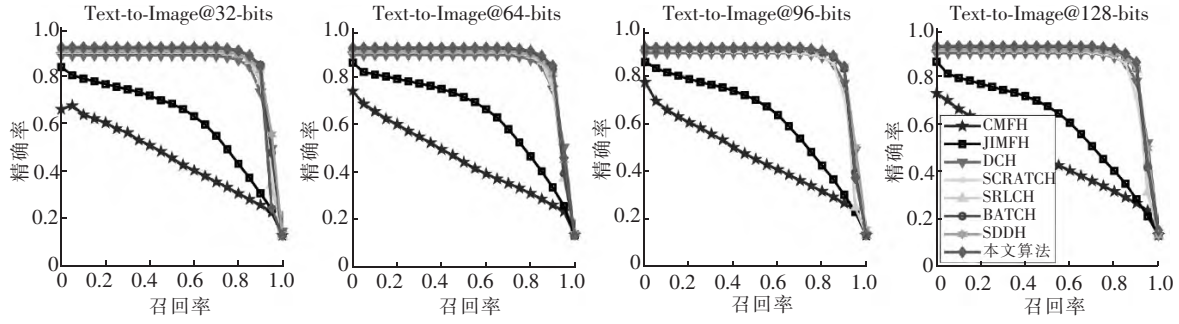

图 3 LabelMe 文本检索图像 PR 曲线图

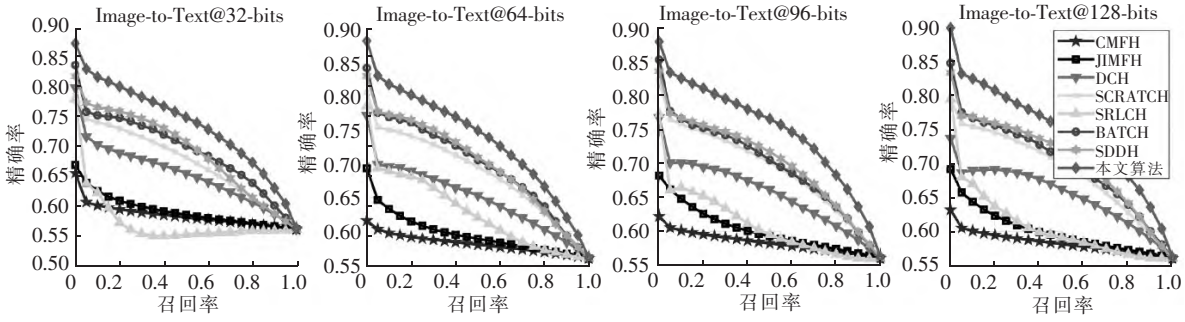

图 4 MIRFlickr 图像检索文本 PR 曲线图

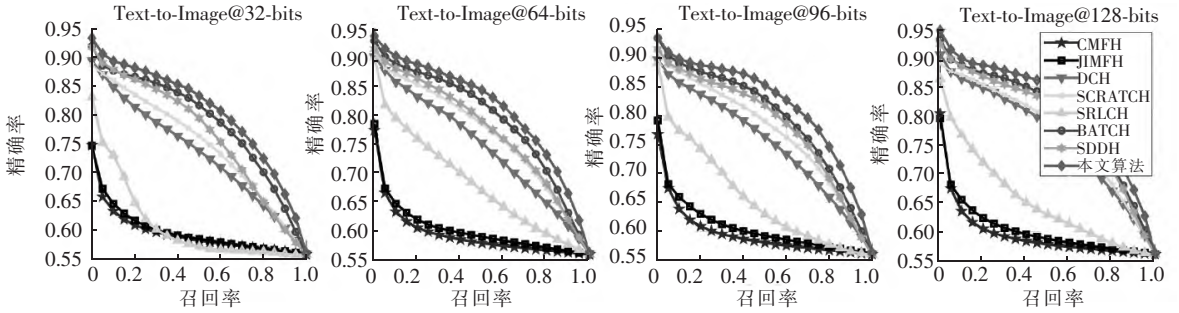

图 5　MIRFlickr 文本检索图像 PR 曲线图

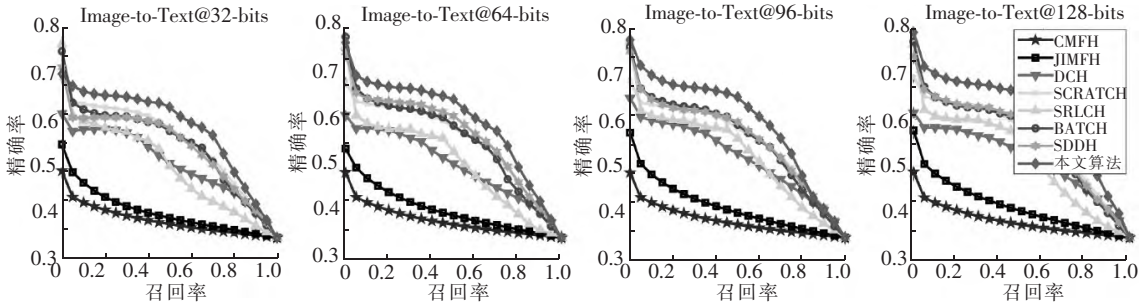

图 6　NUS-WIDE 图像检索文本 PR 曲线图

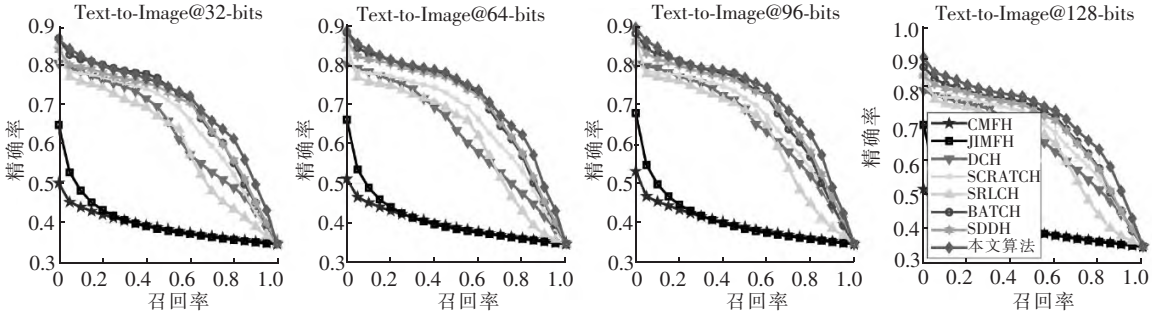

图 7　NUS-WIDE 文本检索图像 PR 曲线图

3.5　收敛性分析

图 8 显示了在 3 个数据集上目标函数值随迭代次数的变化曲线. 如图 8 所示,在 128 位哈希码学习任务中,本文算法在 3 个基准数据集上迭代 5 次后收敛. 这表明本文算法具有快速收敛的特性.

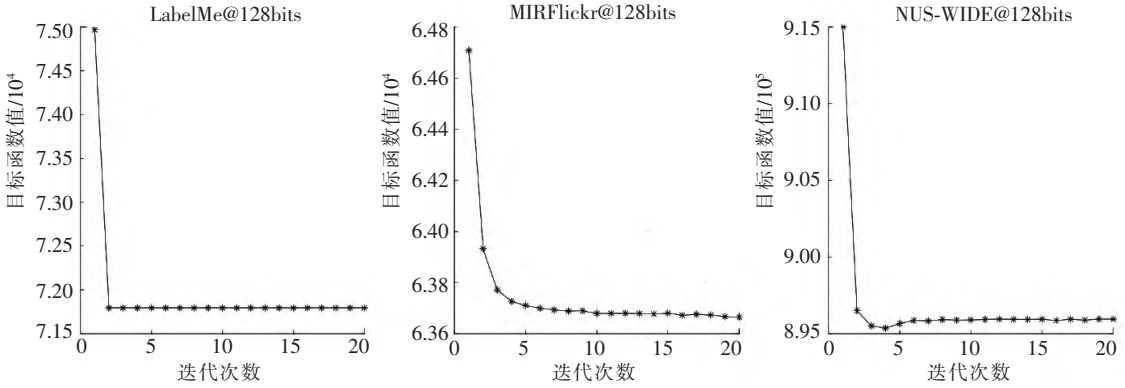

图 8　收敛性分析

3.6　参数敏感性分析

本节进行了参数敏感性分析实验来分析各参数(α,λ,η,δ)的变化对 M_{AP}的影响. 参数 α 控制特征信息在方法中的影响程度,参数 λ 控制特征信息和标签判别性信息融合模块在方法中的影响程度,参数 η 控制特征信息和标签语义信息的侧重学习程度,参数 δ 控制正则化的惩罚程度. 在测试每个参数的同时保持其他参数不变. 4

个参数的变化曲线如图 9 所示. 从图 9 中可观察到:本文方法可以在参数(α,λ,η,δ)较大范围内表现良好和稳定,同时可以观察到 α、λ 和 δ 的敏感性表现基本一致,即在取值为 1 时,M_{AP}的值出现大幅度下降. 这表明特征语义信息、标签语义信息在哈希码的学习过程中的作用是有侧重的. 一般情况,标签语义信息的作用要大于特征语义信息的作用. 对于单个参数来说,当 $\eta = 1$ 时,消除了标签类别语义信息的学习,M_{AP}突然出现骤降,这说明标签所包含的类别语义信息对于哈希码的学习质量非常重要. 对于 δ,可以看出它的值在一定范围内使得模型的拟合程度达到比较好的作用,没有出现过拟合和欠拟合的现象. 对于不同的模型,其 δ 值可能不同.

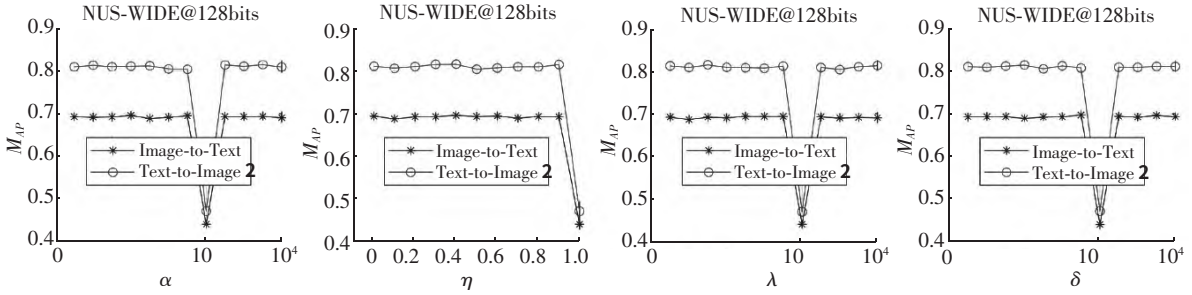

图 9 参数敏感性分析

3.7 时间成本分析

表 5 显示了在 MIRFlickr 数据集上每个方法的训练时间. 如表 5 所示,本文算法的训练时间少于大多数方法的训练时间. 这是因为:本文方法可以同时生成整个哈希码位,而不会随着哈希码长度的增加而产生很大的波动. 可以看出,逐位优化方法(如 DCH)的训练时间相对较长,显示出对哈希码长度的敏感性. 此外,MIRFlickr 的实验设置与真实场景非常相似. 本文算法在 MIRFlickr 数据集上的优异性能表明,本文算法可应用到大规模检索任务中.

表 5 各方法在 MIRFlickr 数据集上的训练时间成本 s

任务	方法	r			
		32	64	96	128
CMFH		6.048 1	7.111 6	7.868 5	9.596 7
JIMFH		97.917 2	111.339 1	123.240 6	141.206 3
DCH		10.783 6	37.485 5	85.026 9	145.332 7
SCRATCH		3.611 9	4.439 2	5.368 8	6.321 6
SRLCH		3.326 0	4.520 9	5.679 0	7.611 1
BATCH		0.762 6	0.924 0	1.266 0	1.503 5
SDDH		1.696 6	1.739 6	2.008 4	2.127 6
本文算法		1.460 9	1.917 1	2.665 9	3.149 6

4 结束语

本文提出了一种标签与样本双语义增强的跨模态哈希检索算法. 该方法利用了特征语义信息、标签语义成对相似性信息和标签类别语义信息来共同监督哈希码的学习. 这不仅实现了对蕴含高级语义信息的多标签数据进行编码的目的,而且能够确保最终学习的哈希码在保持语义相似度的同时还具有较高的差别能力. 本文算法在 3 个广泛使用的数据集上优于目前的先进方法,特别是在多标签数据集上.

5 参考文献

[1] TENG Luyao, TANG Feiyi, ZHENG Zefeng, et al. Kernel based sparse representation learning with global and local low-rank constrain [EB/OL]. [2022-11-01]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9989445.

[2] FANG Xiaozhao, HAN Na, ZHOU Guoxu, et al. Dynamic double classifiers approximation for cross-domain recognition [J]. IEEE Transactions on Cybernetics, 2022, 52(4): 2618-2629.

[3] FANG Xiaozhao, JIANG Kaihang, HAN Na, et al. Average approximate hashing-based double projections learning for cross-modal retrieval [J]. IEEE Transactions on Cybernetics, 2021, 52(11): 11780-11793.

[4] CHEN Zhenduo, LI Chuanxiang, LUO Xin, et al. SCRATCH: a scalable discrete matrix factorization hashing framework for cross-modal retrieval [J]. IEEE Transac-

tions on Circuits and Systems for Video Technology,2020, 30(7):2262-2275.

[5] DING Guiguang,GUO Yuchen,ZHOU Jile. Collective matrix factorization hashing for multimodal data [EB/OL]. [2022-11-01]. https://ieeexplore.ieee.org/document/6909664.

[6] LIN Zijia,DING Guiguang,HU Mingqing,et al. Semantics-preserving hashing for cross-view retrieval [EB/OL]. [2022-03-11]. https://ieeexplore.ieee.org/document/7299011.

[7] LIU Hong,JI Rongrong,WU Yongjian,et al. Cross-modality binary code learning via fusion similarity hashing [EB/OL]. [2022-03-11]. https://ieeexplore.ieee.org/document/8100155/.

[8] LIU Xin,HU Zhikai,LING Haibin,et al. MTFH: a matrix tri-factorization hashing framework for efficient cross-modal retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,43(3):964-981.

[9] QIN Jianyang,FEI Lunke,TENG Shaohua, et al. Discrete semantic matrix factorization hashing for cross-modal retrieval [EB/OL]. [2022-03-15]. https://ieeexplore.ieee.org/document/9413037/.

[10] ZHOU Jile, DING Guiguang, GUO Yuchen. Latent semantic sparse hashing for cross-modal similarity search [EB/OL]. [2022-04-09]. https://dl.acm.org/doi/10.1145/2600428.2609610.

[11] WANG Lu, YANG Jie, ZAREAPOOR M, et al. Cluster-wise unsupervised hashing for cross-modal similarity search [J]. Pattern Recognition, 2021,111(5):107732.

[12] JIN Sheng, YAO Hongxun, ZHOU Qin, et al. Unsupervised discrete hashing with affinity similarity [J]. IEEE Transactions on Image Processing, 2021, 30:6130-6141.

[13] WANG Di, WANG Quan, HE Lihuo, et al. Joint and individual matrix factorization hashing for large-scale cross-modal retrieval [J]. Pattern Recognition,2020,107:107479.

[14] TANG Jun, WANG Ke, SHAO Ling. Supervised matrix factorization hashing for cross-modal retrieval [J]. IEEE Transactions on Image Processing, 2016, 25(7):3157-3166.

[15] QIN Jianyang, FEI Lunke, ZHU Jian, et al. Scalable discriminative discrete hashing for large-scale cross-modal retrieval [EB/OL]. [2022-04-19]. https://ieeexplore.ieee.org/document/9413871.

[16] WANG Yongxin, LUO Xin, NIE Liqiang, et al. BATCH: a scalable asymmetric discrete cross-modal hashing [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(11):3507-3519.

[17] WU Fei, WU Zhiyong, FENG Yujian, et al. Supervised discrete matrix factorization hashing for cross-modal retrieval [EB/OL] [2022-11-15]. https://ieeexplore.ieee.org/document/8691389.

[18] XU Xing, SHEN Fumin, YANG Yang, et al. Learning discriminative binary codes for large-scale cross-modal retrieval [J]. IEEE Transactions on Image Processing, 2017, 26(5):2494-2507.

[19] ZHANG Dongqing, LI Wujun. Large-scale supervised multimodal hashing with semantic correlation maximization [EB/OL]. [2022-11-15]. https://dl.acm.org/doi/10.5555/2892753.2892854.

[20] ZHANG Pengfei, LI Chuanxiang, LIU Mengyuan, et al. Semi-relaxation supervised hashing for cross-modal retrieval [EB/OL]. [2022-11-15]. https://doi.org/10.1145/3123266.3123320.

[21] YAO Tao, YAN Lianshan, MA Yilan, et al. Fast discrete cross-modal hashing with semantic consistency [J]. Neural Networks, 2020, 125(8):142-152.

[22] WANG Di, GAO Xinbo, WANG Xiumei, et al. Label consistent matrix factorization hashing for large-scale cross-modal similarity search [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,41(10):2466-2479.

[23] WANG Song, ZHAO Huan, NAI Kei. Learning a maximized shared latent factor for cross-modal hashing [J]. Knowledge-Based Systems,2021,228(9):107252.

[24] ZHENG Chaoqun, ZHU Lei, LU Xu, et al. Fast discrete collaborative multi-modal hashing for large-scale multimedia retrieval [J]. IEEE Transactions on Knowledge and Data Engineering,2020,32(11):2171-2184.

[25] SHEN Hengtao, LIU Luchen, YANG Yang, et al. Exploiting subspace relation in semantic labels for cross-modal hashing [J]. IEEE Transactions Knowledge and Data Engineering,2021,33(10):3351-3365.

[26] MA Dekui, LIANG Jian, KONG Xiangwei, et al. Discrete cross-modal hashing for efficient multimedia retrieval [EB/OL]. [2022-11-12]. https://ieeexplore.ieee.org/document/7823584.

[27] CHEN Yong, ZHANG Hui, TIAN Zhibao, et al. Enhanced discrete multi-modal hashing: more constraints yet less time to learn [J]. IEEE Transactions on Knowledge and Data Engineering,2022,34(3):1177-1190.

[28] ZHANG Donglin, WU Xiaojun, LIU Zhen, et al. Fast

discrete cross-modal hashing based on label relaxation and matrix factorization [EB/OL]. [2022-11-20]. https://ieeexplore.ieee.org/document/9412497.

[29] 滕少华,郭兰君,张巍,等.一种标签嵌入子空间的跨模态离散哈希学习[J].江西师范大学学报(自然科学版),2021,45(3):305-313.

[30] 李鑫勇,滕少华,张巍,等.语义相似性保持的判别式跨模态哈[J].计算机应用研究,2021,38(11):3359-3365.

[31] 敖宇翔,滕少华,张巍,等.标签局部结构保持的离散哈希方法[J].小型微型计算机系统,2022,43(5):998-1005.

[32] 庄智钧,滕少华,张巍,等.标签松弛回归的跨模态哈希检[J].小型微型计算机系统,2022,43(10):2096-2105.

[33] TENG Shaohua, NING Chengzhen, ZHANG Wei, et al. Fast asymmetric and discrete cross-modal hashing with semantic consistency [EB/OL]. [2023-02-01]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9853632.

[34] LIU Wei, MU Cun, KUMAR S, et al. Discrete graph hashing [EB/OL]. [2022-11-21]. https://dl.acm.org/doi/10.5555/2969033.2969208.

[35] BRYAn C, ANTONIO T, KEVIN P, et al. Labelme: a database and web-based tool for image annotation [J]. International Journal of Computer Vision, 2008, 77(1/3): 157-173.

[36] MARK J, MICHAEL S. The MIRFlickr retrieval evaluation [EB/OL]. [2022-11-18]. https://doi.org/10.1145/1460096.1460104.

[37] TATSENG C, TANG Jinhui, HONG Richang, et al. NUS-WIDE: a real-world Web image database from National University of Singapore [EB/OL]. [2022-10-13]. https://dl.acm.org/doi/10.1145/1646396.1646452.

The Cross-Modal Hash with Tag and Sample Semantic Enhancements

TENG Shaohua[1], HUANG Wenbiao[1], ZHANG Wei[1], TENG Luyao[2]

(1. School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong 510006, China;
2. School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou Guangdong 511483, China)

Abstract: Aiming at the problem that most cross-modal hash methods cannot capture the multi-tag information and the deeper semantic relationship information of feature semantics, a cross-modal retrieval framework with bilingual enhancement of tag and sample is proposed. The framework first decomposes different high-dimensional modal data into low-dimensional shared feature semantic space. Secondly, the hash code learning function that relaxes the variables into the tag semantic constraints is introduced to strengthen the sample semantic similarity hash code learning by minimizing the tag pair distance, which not only maintains the relationship between the cross-modal corresponding sample semantics, strengthens the tag semantic learning of the hash code, but also solves the problem of solving the real symmetric matrix and the convergence of the algorithm. Thirdly, further apply sample feature semantics and tag semantics to enhance the semantic learning of hash codes. Finally, the experimental results on three commonly used data sets show that this method is superior to the current advanced methods.

Key words: tag and sample semantic enhancements; cross-modal retrieval; tag semantics

(责任编辑:冉小晓)