有监督鉴别哈希跨模态检索

朱治兰[1] 荆晓远[1*] 董西伟[1,2] 吴 飞[1]

[1](南京邮电大学自动化学院 江苏 南京 210023)

[2](九江学院信息科学与技术学院 江西 九江 332005)

摘 要 随着大数据时代的到来,利用哈希方法实现对异质多模态数据的快速跨模态检索受到越来越多的关注。为了获取更好的跨模态检索性能,提出有监督鉴别跨模态哈希算法。利用对象的标签信息对所要生成的哈希码进行约束。算法中的线性分类项和图拉普拉斯算子项分别用于提升哈希码鉴别能力和保留模态间相似性。对算法的目标函数利用迭代法进行求解。该算法在两个基准数据集的实验结果展现出优于目前最前沿的跨模态哈希检索方法。

关键词 跨模态检索 哈希 标签信息 线性分类项 图拉普拉斯项

中图分类号 TP3 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.04.035

SUPERVISED DISCRIMINATIVE CROSS-MODAL HASHING

Zhu Zhilan[1] Jing Xiaoyuan[1*] Dong Xiwei[1,2] Wu Fei[1]

[1](*College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China*)

[2](*School of Information Science and Technology, Jiujiang University, Jiujiang 332005, Jiangxi, China*)

Abstract With the advent of the era of big data, applying Hash methods for heterogeneous multimodal data to achieve fast cross-modal retrieval is receiving more and more attention. In order to obtain better cross-modal retrieval performance, we proposed supervised discriminative cross-modal hashing(SDCH). We used label information of objects to constrain the hash code to be generated. The linear classification term and the graph Laplacian term in the algorithm were used to improve the discriminating ability of hash codes and preserve inter-modal similarity, respectively. The objective function of the algorithm was solved in an iterative method. Related experiments have been conducted in two benchmark databases. Experimental results show that the algorithm is superior to state-of-the-art cross-modal Hash methods.

Keywords Cross-modal retrieval Hash Label information Linear classification Graph Laplacian term

0 引 言

近几十年来,互联网多媒体数据的爆炸性增长,使得跨媒体数据检索需求增长,并且促进了复杂多模态检索技术的发展。现如今,多媒体数据往往来自不同的互联网多媒体平台以及不同的数据资源。这些数据经常共同出现且被用来描述同一物体和事件,因此跨模态检索在实际应用中已经成为必要。例如,人们经常利用图片来检索相关的文本文献,或者用文本来检索相关的图片内容。但是,由于多模态数据属于不同的特征空间,这种异质的特征被认为是跨模态检索最大的挑战。

为了消除不同模态特征之间的异质性,最近有很多研究把关注点放在潜在子空间的学习上。研究的关键点在于如何通过学习得到一个共同的语义子空间,使得不同模态数据之间的异质性得到消除,进而使得这些特征在这个学习得到的语义子空间中能被直接相互匹配。然而,由于忽视了特征维度的可伸缩性,在解决大规模数据的多模态检索时,这些方法受到了限制。

收稿日期:2018-10-29。国家自然科学基金项目(61702280);江苏省研究生创新工程项目(CXLX11_0418)。**朱治兰**,硕士生,主研领域:模式识别,跨模态检索。**荆晓远**,教授。**董西伟**,讲师。**吴飞**,讲师。

此时出现了大量的哈希方法。之前大多数的哈希方法目的在于生成多模态数据的哈希码,生成哈希码的方法可以分为两大类,一类是依赖训练数据的,另一类是独立于训练数据的。其中有一种不依赖训练数据的著名算法是局部敏感哈希算法[1],它随机地选取线性投影矩阵作为哈希函数。相比独立于训练数据的哈希算法,依赖训练数据的哈希算法提供了一种更加可靠的投影机制,从而能够得到更加简洁以及有鉴别力度的哈希码。

从有无利用标签信息的角度来看,现有的跨模态哈希算法又被划分成有监督的哈希算法[2-3]和无监督的哈希算法[4-6]。无监督的方法一般是利用训练数据之间的相关性学习一种能够将多模态特征转化为哈希码的投影机制,而有监督的方法则是将训练数据的标签信息作为一种语义约束,从而得到更加合适的哈希码。前人的实验表明有监督的哈希算法在跨模态检索中取得的效果通常比无监督的哈希算法取得的效果要好。

对有监督跨模态哈希学习算法,一些研究者对模态间和模态内的相似性保留问题进行了研究。而且,大部分的有监督跨模态哈希具有相同的特性,即通过标签信息保留模态间和模态内的相似性来学习哈希码。然而,这一特性的缺点是削弱了学习到的哈希码的鉴别力度。第一,保留模态间和模态内的相似性不能保证学习到的哈希码在语义上的鉴别力。第二,计算给定标签下的成对相似性不可避免地会造成额外的存储和计算损耗。

为了解决上述问题,本文提出了一种新的跨模态哈希算法,即:有监督鉴别跨模态哈希算法 SDCH(Supervised Discriminative Cross-modal Hashing)。与现有的跨模态哈希方法[6-7]相比,SDCH 不仅可以节省存储空间还可以减少在线检索时间。图 1 描述了整个工作流程。SDCH 利用模态间的标签信息作为主要的约束条件,来保留模态间数据的相似性,结合线性分类器来学习得到多模态数据的统一的有鉴别力的哈希码。

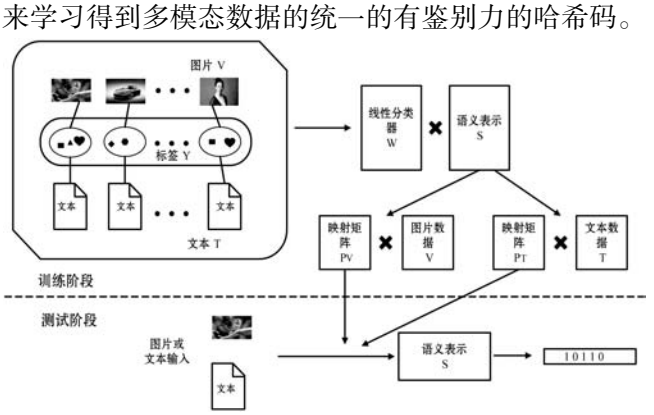

图 1 算法流程

本文的主要贡献总结为以下几个方面:

1)将模态间相似性的保留融合到分类器框架中,从而学习到既能保证相似性又能体现鉴别力的统一哈希码。

2)仅使用模态间的标签信息作为主要的约束条件,与传统的相似性保留方法相比减少了计算时间,节省了存储消耗,同时还能保留模间同标签数据之间的相似性。

3)本文在两个数据集上做了实验来评估 SDCH 方法,实验结果显示 SDCH 较前沿方法具有一定的竞争性。

1 相关工作

1.1 潜在子空间的学习

对于跨模态检索问题,最近许多研究都把关注点放在潜在子空间学习上。其中,典型相关性分析 CCA(Canonical Correlation Analysis)是目前最流行跨模态检索方法之一,它主要学习一对能够使得两种模态数据投影到共同的潜在空间上的投影变换,而投影后得到的数据能够直接匹配,且能够最大程度地反映出两种模态数据之间的关系。CCA 被广泛地研究并延伸出许多相关算法。例如,Rasiwasia 等[9]提出的在跨模态检索方法能够学习多模态数据的最大相关子空间。其他经典的算法还有双线性分离样式模型[10]和广义耦合字典学习[11],它们和 CCA 非常相似都是学习共同的潜在子空间来进行跨模态检索工作。除了以上经典的方法以外,还有一些子空间学习方法利用分类标签信息来改善检索性能,例如,Sharma 等[12]提出的一种普通的多模态分析算法,利用标签信息学习有鉴别的潜在子空间;Gong 等[13]提出的多模态 CCA 框架,基于标签的语义信息将图像和文本两种模态的数据联系在一起进行跨模态检索;再有深度跨模态模型,比如深度 CCA[14]、多模态自编码[15]以及多模态限制玻尔兹曼机[16]等都是用来解决模态检索问题而提出的算法。这些算法的目的都是希望学习到能够更好地保留原始数据特征的子空间。

1.2 哈希模型

随着高维度跨模态数据的爆炸性增长,最近邻搜索花费的代价越来越高。为了解决这个问题,基于哈希的一系列算法被提出,比如文献[17-20],在大规模多模态数据检索领域上得到了广泛的关注。文献[4]中提出的算法第一次将哈希思想运用到多模态检索问题上。不过这个算法的缺陷在于它只考虑到了模

态内数据的相关性而忽略了模态间数据的相关性。为了解决这一问题,文献[21]中通过最小化相似数据哈希码之间的的距离并最大化不相关数据哈希码之间的距离来生成跨模态检索的哈希码。更近一步地,Wu等[22]提出了稀疏多模态哈希算法,这一算法通过联合的多模态字典学习获取不同模态数据的稀疏哈希码从而解决跨模态检索问题。为了更好地利用多模态数据的标签信息,Tang 等[23]提出了有监督的矩阵分解哈希 SMFH(Supervised Matrix Factorization Hashing for Cross-Modal Retrieval),该算法利用集体矩阵分解技术得到统一的哈希码,并且结合不同模态数据的标签一致性以及模态内数据的局部几何一致性使得得到的哈希码具有更好的鉴别力。

然而,这种通过标签信息来保留哈希码相似性的有监督学习算法没能通过给定的标签信息挖掘分类信息,从而使得学习得到的哈希码缺乏鉴别力。而且,这种相似性的保留,会占用更大的存储空间以及消耗更多的检索时间。为此,本文提出了一种新的算法 SDCH,该算法利用相同对象不同模态间的数据具有相同标签信息这一特性来保留成对哈希码的相似性,同时又利用标签信息构造能使哈希码所表示的数据被正确分类的线性分类器来保证哈希码的鉴别力。

2　算法设计

本节将描述所提算法 SDCH 的细节。为了简洁地阐述该算法,以两种模态(图像和文本)哈希码的学习为例。不失一般性,它可以延伸到多模态数据的学习上。

首先,假设本文有两种模态的训练数据 $\boldsymbol{V}=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$ 和 $\boldsymbol{T}=\{\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n\}$,它们分别是同一对象的两种表示模态,这里的 n 为训练样本的个数。每个对象都由对应的图像和文本共同组成。对于第 i 个对象,$\boldsymbol{v}_i\in\boldsymbol{R}^{d_1}$ 是 d_1 维度的图像特征向量,$\boldsymbol{t}_i\in\boldsymbol{R}^{d_2}$ 是 d_2 维度的文本特征向量(在大部分情况下 $d_1\neq d_2$)。不失一般性,本文对跨模态数据进行中心化处理,即令 $\sum_i^n \boldsymbol{v}_i=0$ 以及 $\sum_i^n \boldsymbol{t}_i=0$。

算法 SDCH 的目的是分别为图像和文本找到能够使得特征向量转化为统一的哈希码的哈希函数,即 $f(v):R^{d_1}\rightarrow\{-1,1\}^k$ 以及 $g(t):R^{d_2}\rightarrow\{-1,1\}^k$,这里的 k 指代的是哈希码的长度。为了得到有意义的哈希码,本文借鉴文献[3]中使用的方法将异质数据投影到同一汉明空间中,同时假设模态间同标签数据具有相同哈希码。这样图像数据和文本数据就能在被投影到汉明空间的同时保留原始数据的语义信息。

2.1　学习有鉴别的哈希码

本文考虑来自两种模态的特征 $\boldsymbol{v}_i$ 和 $\boldsymbol{t}_i$ 具有相同的哈希码表示 s_i。本文希望这里的 s_i 能够消除来自两种模态数据原始特征的语义鸿沟。正如图 1 所描述的,原始特征被投影到一个共同的汉明空间中。

因此对于跨模态数据,本文分别通过两种线性变换投影原始图像和文本特征到汉明空间:

$$\boldsymbol{S}_V=\boldsymbol{P}_V\boldsymbol{V} \tag{1}$$

$$\boldsymbol{S}_T=\boldsymbol{P}_T\boldsymbol{T} \tag{2}$$

式中:$\boldsymbol{P}_V\in\boldsymbol{R}^{k\times d_1}$,$\boldsymbol{P}_T\in\boldsymbol{R}^{k\times d_2}$。

基于同对象不同模态的数据具有相同语义表示这一假设,本文通过最小化以下函数来求解两个线性变换矩阵:

$$O_{lp}(\boldsymbol{P}_V,\boldsymbol{P}_T)=\|\boldsymbol{S}-\boldsymbol{S}_V\|_F^2+\|\boldsymbol{S}-\boldsymbol{S}_T\|_F^2=\|\boldsymbol{S}-\boldsymbol{P}_V\boldsymbol{V}\|_F^2+\|\boldsymbol{S}-\boldsymbol{P}_T\boldsymbol{T}\|_F^2 \tag{3}$$

式中:$\|\cdot\|_F$ 表示矩阵的 Frobenius 范数。

此外,原始多模态数据特征还可以反映分类信息,为了让本文得到的哈希码也能够反映这一特性,那么得到的哈希码就能够通过它们的原始标签被分类[8]。因此假设给定第 i 个目标的标签向量 $\boldsymbol{y}_i$,本文就可以用线性分类器 $\boldsymbol{W}\in\boldsymbol{R}^{k\times c}$ 来预测哈希码的标签向量,即:

$$\boldsymbol{Y}=\boldsymbol{W}^{\mathrm{T}}S \tag{4}$$

线性分类器 $\boldsymbol{W}$ 可以通过最小化以下函数来求解:

$$O_{mf}(\boldsymbol{W},\boldsymbol{S})=\|\boldsymbol{Y}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}\|_F^2 \tag{5}$$

式中:$\boldsymbol{Y}\in\boldsymbol{R}^{c\times n}$ 表示 n 个标签向量组成的矩阵。

2.2　拉普拉斯项

为了利用标签信息,本文为双模态数据之间的标签一致性建模,并且将图像和文本两种模态数据之间的语义类同度量为:

$$C_{ij}=\begin{cases}1 & \text{if } v_i \text{ and } t_j \text{ have the same catagory}\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

为了保留两种模态数据在汉明空间中的标签一致性,本文可以最小化以下函数:

$$O_c(\boldsymbol{S})=\sum_{i,j=1}^{n}c_{ij}\|s_i-s_j\|^2 \tag{7}$$

式中:$\|\cdot\|$ 为二范数。

通过一系列的代数运算,可以将式(7)重新生成为:

$$O_c(\boldsymbol{S})=\sum_{i,j=1}^{n}c_{ij}\|s_i-s_j\|^2=2tr(\boldsymbol{S}(\boldsymbol{D}-\boldsymbol{C})\boldsymbol{S}^{\mathrm{T}})=2tr(\boldsymbol{S}\boldsymbol{L}\boldsymbol{S}^{\mathrm{T}}) \tag{8}$$

式中：$\boldsymbol{C}\in\boldsymbol{R}^{n\times n}$为相似性矩阵；$\boldsymbol{D}\in R^{n\times n}$是对角矩阵，其中对角线元素为 $\boldsymbol{C}$ 矩阵列之和，即：$D_{ii}=\sum_j C_{ij}$。因此 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{C}$ 为拉普拉斯矩阵的形式。

2.3 目标函数

完整的目标函数由以下几个部分共同组成，分别是式(5)有鉴别项 O_{mf}、式(3)线性变换项 O_{lp}、式(8)拉普拉斯项 O_c 以及 Frobenius 范数项，具体如下式所示：

$$\begin{aligned}&\min_{\boldsymbol{W},\boldsymbol{P}_V,\boldsymbol{P}_T,\boldsymbol{S}} O(\boldsymbol{W},\boldsymbol{P}_V,\boldsymbol{P}_T,\boldsymbol{S})=\\&O_{mf}+O_{lp}+O_c+\lambda\|\boldsymbol{W}\|_F^2=\\&\|\boldsymbol{Y}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}\|_F^2+\mu_V\|\boldsymbol{S}-\boldsymbol{P}_V\boldsymbol{V}\|_F^2+\mu_T\|\boldsymbol{S}-\boldsymbol{P}_{\mathrm{T}}\boldsymbol{T}\|_F^2+\\&\gamma tr(\boldsymbol{S}\boldsymbol{L}\boldsymbol{S}^{\mathrm{T}})+\lambda\|\boldsymbol{W}\|_F^2\end{aligned}\tag{9}$$

这里的正则项 $\|\cdot\|_F^2$ 起到防止过拟合的作用。$\lambda=0.01$ 为正则系数。μ_V、μ_T 为惩罚因子，其值设置为 10^{-5}。$\gamma=1$ 为折衷参数。

2.4 迭代优化

式(9)所示的最优化问题是拥有四个矩阵型变量的非凸函数，所以解决他是具有一定困难的。不过，当固定其他三个变量只求解其中一个变量的情况下，它又是凸的。因此，关于这一最优化问题的求解就可以被总结成如下三步迭代算法：

第一步：固定 $\boldsymbol{S}$ 和 $\boldsymbol{W}$ 求 $\boldsymbol{P}_V$ 和 $\boldsymbol{P}_T$。当固定 $\boldsymbol{S}$ 和 $\boldsymbol{W}$ 后，式(9)转化为如下的优化问题：

$$\min_{\boldsymbol{P}_V,\boldsymbol{P}_T} O(\boldsymbol{P}_V,\boldsymbol{P}_T)=\mu_V\|\boldsymbol{S}-\boldsymbol{P}_V\boldsymbol{V}\|_F^2+\mu_T\|\boldsymbol{S}-\boldsymbol{P}_T\boldsymbol{T}\|_F^2\tag{10}$$

令 $\frac{\partial O}{\partial \boldsymbol{P}_V}=0$，$\frac{\partial O}{\partial \boldsymbol{P}_T}=0$，进一步推导可得：

$$\boldsymbol{P}_V=\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}+\frac{\boldsymbol{I}}{\mu_V}\right)^{-1}\tag{11}$$

$$\boldsymbol{P}_T=\boldsymbol{S}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{T}\boldsymbol{T}^{\mathrm{T}}+\frac{\boldsymbol{I}}{\mu_T}\right)^{-1}\tag{12}$$

式中：$\boldsymbol{I}$ 表示单位矩阵，$(\cdot)^{-1}$ 表示所求矩阵的逆运算。

第二步：固定 $\boldsymbol{P}_V$ 和 $\boldsymbol{P}_T$ 以及 $\boldsymbol{S}$ 求 $\boldsymbol{W}$。当固定 $\boldsymbol{P}_V$ 和 $\boldsymbol{P}_T$ 以及 $\boldsymbol{S}$ 后，式(9)转化为如下的优化问题：

$$\min_{\boldsymbol{W}}\mu_V\|\boldsymbol{Y}-\boldsymbol{W}^{\mathrm{T}}\|_F^2+\lambda\|\boldsymbol{W}\|_F^2\tag{13}$$

令 $\frac{\partial O}{\partial \boldsymbol{W}}=0$，进一步推导可得：

$$\boldsymbol{W}=(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{S}\boldsymbol{Y}^{\mathrm{T}}\tag{14}$$

第三步：固定 $\boldsymbol{P}_V$ 和 $\boldsymbol{P}_T$ 以及 $\boldsymbol{W}$ 求 $\boldsymbol{S}$。当固定 $\boldsymbol{P}_V$ 和 $\boldsymbol{P}_T$ 以及 $\boldsymbol{W}$ 后，式(9)转化为如下的优化问题：

$$\min_{\boldsymbol{S}}\|\boldsymbol{Y}-\boldsymbol{W}^{\mathrm{T}}\|_F^2+\mu_V\|\boldsymbol{S}-\boldsymbol{P}_V\|_F^2+\mu_T\|\boldsymbol{S}-\boldsymbol{P}_T\|_F^2+\gamma tr(\boldsymbol{S}\boldsymbol{L}\boldsymbol{S}^{\mathrm{T}})\tag{15}$$

令 $\frac{\partial O}{\partial \boldsymbol{S}}=0$ 整理得：

$$\boldsymbol{A}\boldsymbol{S}+\boldsymbol{S}\boldsymbol{B}+\boldsymbol{E}=0\tag{16}$$

式中：

$$\boldsymbol{A}=2(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}+(\mu_V+\mu_{\mathrm{T}})\boldsymbol{I})$$

$$\boldsymbol{B}=\boldsymbol{L}+\boldsymbol{L}^{\mathrm{T}}$$

$$\boldsymbol{E}=-2(\boldsymbol{W}\boldsymbol{T}+\mu_V\boldsymbol{P}_V\boldsymbol{V}+\mu_T\boldsymbol{P}_T\boldsymbol{T})$$

值得注意的是式(16)是希尔维斯特方程，可以用 MATLAB 中的李雅普诺夫函数求解。

SDCH 算法可以被概括成算法 1。当一个新的检索条目 x_q 被输入，SDCH 首先会利用式(1)或者式(2)生成语义表示 s_q 然后本文会根据符号函数 $H^{x_q}=sign(s_q)$ 生成想要的哈希码。

算法 1 有监督鉴别哈希跨模态检索

输入：图片特征矩阵 $\boldsymbol{V}$，文本特征矩阵 $\boldsymbol{T}$，两种模态的语义标签 Y，系数 λ，μ_V，μ_T，γ，以及哈希码长度 k。

输出：哈希码 H，投影矩阵 $\boldsymbol{P}_V$，$\boldsymbol{P}_T$。

1：中心化 $\boldsymbol{V}$，$\boldsymbol{T}$，根据式(7)和式(8)构造拉氏矩阵

2：随机初始化 $\boldsymbol{P}_V$，$\boldsymbol{P}_T$，$\boldsymbol{S}$，$\boldsymbol{W}$。

3：重复步骤 4－6

4：固定 $\boldsymbol{S}$、$\boldsymbol{W}$，根据式(11)和式(12)更新 $\boldsymbol{P}_V$、$\boldsymbol{P}_T$。

5：固定 $\boldsymbol{P}_V$、$\boldsymbol{P}_T$、$\boldsymbol{S}$，根据式(14)更新 $\boldsymbol{W}$。

6：固定 $\boldsymbol{P}_V$、$\boldsymbol{P}_T$、$\boldsymbol{W}$，根据式(16)更新 $\boldsymbol{S}$。

7：直到收敛

8：$H=sign(\boldsymbol{S})$

3 实 验

本文将在 Wiki，NUS_WIDE 两个数据库上进行实验，而后将实验结果和最前沿方法生成的结果进行对比。

3.1 实验设置

3.1.1 数据集

Wiki：它是从维基百科中搜集来的 2 866 个图像-文本对的集合。其中图像是由 128 维度的 SIFT 特征表示的，而文本则是由 10 维度的主题向量特征构成的。本文所用的 Wiki 数据集包含了 10 个语义分类，并且随机抽取其中的 2 173 个数据对作为训练集，将剩余的 693 个数据对作为测试集。

NUS_WIDE：本文实验中所用到的 NUS_WIDE 数据集是从 Flickr 中下载到的网页图像集合。原始数据集是包含了 81 个主题的且都有标注信息的 269 648 幅图像数据的集合。这样每幅图像和它对应的标注信息就构成了一个图像-文本对。本文从中挑选包含 186 577 幅前十类的图片作为实验数据。其中，图像数据是由 500 维度的视觉词袋 SIFT 直方图表示的，基于

top-1000 标注信息生成文本的词袋特征向量表示。对于所挑选的数据集,本文从中随机地挑选 5 000 图像文本对作为训练集,然后在剩余数据中再随机挑选 1 866 图像文本对作为测试集。

3.1.2　对比算法

本文将 SDCH 算法与五个最前沿的算法典型相关性分析 CCA(Canonical Correlation Analysis)[23]、集体矩阵分解哈希 CMFH(Collective Matrix Factorization Hashing)[3]、交叉视图哈希模型 CVH(Cross-View Hashing Model)[21]、潜在语义稀疏哈希 LSSH(Latent Semantic Sparse Hashing)[2] 和有监督的矩阵分解哈希 SMFH(Supervised Matrix Factorization Hashing)[24]作对比。

其中,CCA 算法将双模态数据投影到一个能使数据之间的相关性最大化的同一空间中;CMFH 通过集体矩阵分解发现不同模态数据的潜在因子模型,从而学习统一的哈希码;CVH 通过解决广义特征值问题最小化数据对的加权平均多视图的 l2 范式距离;LSSH 假设同一对象不同模态数据之间具有完全相同的哈希码,并首次将稀释编码和矩阵分解结合到一起来分别学习文本和图像的语义特征,为了缩小语义鸿沟继而将其投影到抽象空间中;SMFH 利用矩阵分解技术的同时,考虑不同模态数据之间的标签一致性以及同模态数据之间的局部几何结构的一致性。

3.1.3　评估度量

本文将进行两种跨模态检索任务,一种是“Img to Txt”,即利用图像来搜索相关的文本。另一种是“Txt to Img”,即利用文本来搜索相关的图片。为了评估跨模态检索的性能,本文引入平均精度均值 mAP(mean Average Precision)这一度量标准:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}(q_i)$$

式中:q_i是一条检索输入且 N 是检索条目输入总数。平均精度 AP(Average Precision)的计算公式如下:

$$AP(q) = \frac{1}{T}\sum_{r=1}^{R} P_q(r)\xi(r)$$

式中:T 是检索集中所有的相关实体的个数,$P_q(r)$是按照相关度排名后的前 r 个实体的精度,$\xi(r)$是一个指示函数,当第 r 个被检索到的实体与检索内容相关则其值为 1 否则为 0。

本文还学习了 Wiki 数据集上的表现曲线,即精度-召回曲线(PR-Curves)。精度-召回曲线是精度值关于召回值的函数,它被广泛地运用到跨模态检索性能的评估上。因为评估数据是随机选取的,所以本文取 10 次实验的平均值作为最后的结果。

3.2　结果及分析

表 1 和表 2 展示了本文提出的 SDCH 算法和五个对比算法的 mAP 值。通过观察表 1 和表 2 可以看出,与对比算法相比,本文所提出的 SDCH 算法在不同哈希码长度下都具有较好的 mAP 值。这说明本文提出的 SDCH 算法能够挖掘到更多的鉴别信息来提升跨模态检索性能,这得益于标签信息的利用保留了跨模态数据之间的相似性也得益于线性分类器思想的应用提高了哈希码的鉴别力。通过观察表 1 和表 2 还可以看到,在哈希码比较短的 16 位时,SDCH 算法相较于 SMFH 算法在 mAP 值上也有很大的优势,这进一步表明了 SDCH 算法在实质上作了改善。此外,结果还表明,随着哈希码长度的增加,SDCH 的性能就越好。对比 SDCH 算法的“Img to Txt”和“Txt to Img”两个任务还发现“Txt to Img”任务的检索效果总是优于“Img to Txt”任务的检索效果,且对于数据尺度较大的 NUS-WIDE 数据集而言“Txt to Img”任务的检索效果的得到了显著的提升。

表 1　Wiki 数据集上的 mAP 值

任务	方法	哈希码长度			
		16 位	32 位	64 位	128 位
Img to Txt	CCA	0.171 4	0.154 7	0.146 5	0.128 6
	CMFH	0.211 6	0.226 0	0.239 4	0.241 4
	CVH	0.171 8	0.157 3	0.150 8	0.135 1
	LSSH	0.216 2	0.224 8	0.225 5	0.219 0
	SMFH	0.269 9	0.283 5	0.292 0	0.298 1
	SDCH	0.298 8	0.324 1	0.350 9	0.359 9
Txt to Img	CCA	0.159 0	0.141 5	0.130 0	0.115 1
	CMFH	0.483 0	0.523 5	0.536 1	0.531 8
	CVH	0.150 5	0.134 7	0.121 4	0.113 6
	LSSH	0.501 0	0.524 6	0.532 4	0.538 7
	SMFH	0.605 1	0.625 7	0.635 7	0.642 8
	SDCH	0.696 4	0.715 5	0.721 2	0.728 8

表 2　NUS_WIDE 数据集上 mAP 值

任务	方法	哈希码长度			
		16 位	32 位	64 位	128 位
Img to Txt	CCA	0.360 6	0.353 8	0.349 9	0.347 2
	CMFH	0.372 8	0.379 3	0.378 8	0.377 1
	CVH	0.373 8	0.362 7	0.355 5	0.350 6
	LSSH	0.383 8	0.383 7	0.390 3	0.386 4
	SMFH	0.455 3	0.462 3	0.465 8	0.468 0
	SDCH	0.538 9	0.575 4	0.575 3	0.578 9

续表 2

任务	方法	哈希码长度			
		16 位	32 位	64 位	128 位
Txt to Img	CCA	0.359 8	0.355 3	0.350 2	0.348 2
	CMFH	0.373 4	0.378 5	0.382 3	0.381 9
	CVH	0.375 7	0.363 6	0.356 7	0.356 2
	LSSH	0.415 0	0.411 4	0.416 5	0.413 2
	SMFH	0.503 3	0.505 6	0.506 5	0.507 9
	SDCH	0.700 6	0.708 7	0.724 3	0.691 4

图 2 和图 3 显示了 SDCH 算法和五个对比算法的精度-召回曲线。通过观察发现 SDCH 算法表现优于其他的对比算法,这与用 mAP 对算法性能进行评价的结果一致。通过观察还发现,对于 CCA 和 CVH 算法而言,精度-召回值更像是随机出现的,所以对于这两个算法分析它的精度-召回曲线几乎是没有意义的。

最后,通过观察还可以看到,SDCH 在 16 位哈希码的条件下进行检索时的效果甚至超过了对比算法在更长的哈希码下检索的效果,这也就充分体现出了本文所提算法在性能上的优越性。

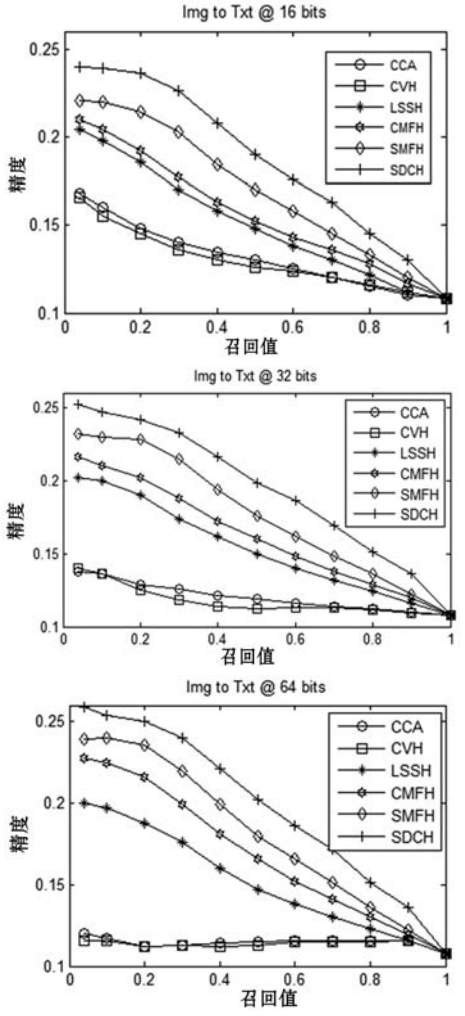

图 2 Wiki 数据集在不同哈希码长度上的精度-召回曲线(Img to Txt)

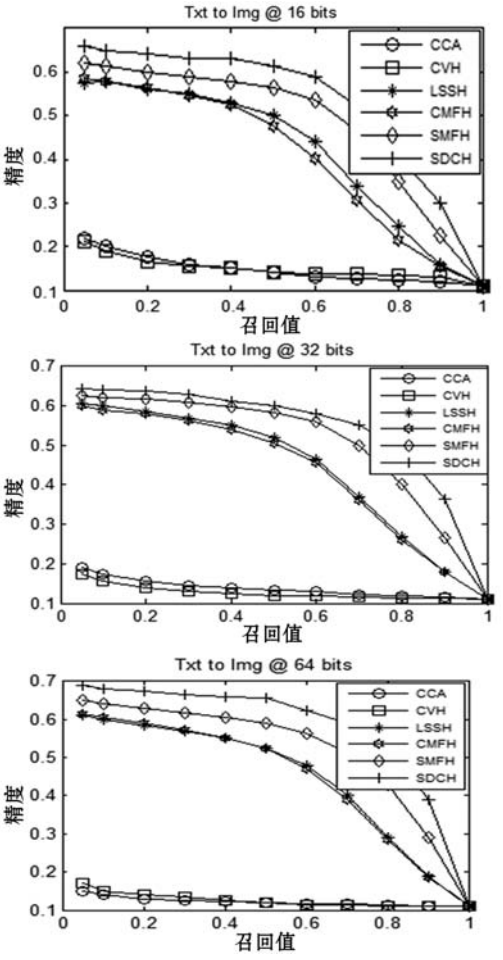

图 3 Wiki 数据集在不同哈希码长度上的精度-召回曲线(Txt to img)

4 结 语

本文提出了一种新的跨模态检索方法,即有监督鉴别跨模态哈希算法。为了得到有鉴别力的哈希码,本文将哈希码的学习嵌入到线性分类器的框架中,其中线性分类器部分公式的形成利用了标签信息的监督原理。此外为了不损坏模态间数据的相似性,本文利用模态间数据的标签一致性作为相似性的约束。

本文在两个常用的数据集 Wiki 和 NUS_WIDE 上进行了实验来验证本文所提算法的有效性。本文将 SDCH 方法的实验结果和几种最前沿跨模态哈希检索算法的实验结果进行了对比和分析评估,结果显示所提算法 SDCH 能够取得更好的跨模态检索性能。

参 考 文 献

[1] Liu L, Yu M, Shao L. Multiview alignment hashing for efficient image search[J]. IEEE Transactions on Image Processing, 2015, 24(3): 956-966.

[2] Zhou J, Ding G, Guo Y. Latent semantic sparse hashing for cross-modal similarity search[C]//Proceedings of the 37th

International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM, 2014: 415 - 424.

[3] Ding G, Guo Y, Zhou J. Collective matrix factorization hashing for multimodal data[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2014: 2075 - 2082.

[4] Bronstein M M, Bronstein A M, Michel F, et al. Data fusion through cross-modality metric learning using similarity-sensitive hashing[C]//Computer Vision and Pattern Recognition, 2010 IEEE Conference. IEEE, 2010: 3594 - 3601.

[5] Zhang D, Li W J. Large-Scale Supervised Multimodal Hashing with Semantic Correlation Maximization[C]//The Association for the Advancement of Artificial Intelligence. AAAI, 2014, 1(2): 7 - 18.

[6] Zhen Y, Yeung D Y. Co-regularized hashing for multimodal data[C]//Advances in neural information processing systems. NIPS, 2012: 1376 - 1384.

[7] Irie G, Arai H, Taniguchi Y. Alternating co-quantization for cross-modal hashing[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015: 1886 - 1894.

[8] Xu X, Shen F, Yang Y, et al. Learning discriminative binary codes for large-scale cross-modal retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2494 - 2507.

[9] Rasiwasia N, Costa Pereira J, Coviello E, et al. A new approach to cross-modal multimedia retrieval[C]//Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010: 251 - 260.

[10] Luo Y, Liu T, Tao D, Xu C. Multiview matrix completion for multilabel image classification[J]. IEEE Transactions on Image Processing, 2015, 24(8): 2355 - 2368.

[11] Mandal D, Biswas S. Generalized coupled dictionary learning approach with applications to cross-modal matching[J]. IEEE Transactions on Image Processing, 2016, 25(8): 3826 - 3837.

[12] Sharma A, Kumar A, Daume H, et al. Generalized multiview analysis: A discriminative latent space[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2160 - 2167.

[13] Gong Y, Ke Q, Isard M, et al. A multi-view embedding space for modeling internet images, tags and their semantics [J]. International journal of computer vision, 2014, 106(2): 210 - 233.

[14] Andrew G, Arora R, Bilmes J, et al. Deep canonical correlation analysis[C]//International Conference on Machine Learning. ICML, 2013: 1247 - 1255.

[15] Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning[C]//Proceedings of the 28th international conference on machine learning. ICML, 2011: 689 - 696.

[16] Srivastava N, Salakhutdinov R R. Multimodal learning with deep boltzmann machines[C]//Advances in neural information processing systems. NIPS, 2012: 2222 - 2230.

[17] Liu T, Tao D. Classification with noisy labels by importance reweighting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(3): 447 - 461.

[18] Jiang R, Qiao H, Zhang B. Speeding up graph regularized sparse coding by dual gradient ascent[J]. IEEE signal processing letters, 2015, 22(3): 313 - 317.

[19] Xu C, Liu T, Tao D. Local rademacher complexity for multi-label learning[J]. IEEE Transactions on Image Processing, 2016, 25(3): 1495 - 1507.

[20] Wang J, Kumar, Chang S F. Semi-supervised hashing for scalable image retrieval[C]//IEEE Conference on Computer Vision & Pattern Recognition. DBLP, 2010: 1654 - 1661.

[21] Kumar S, Udupa R. Learning Hash Functions for Cross-View Similarity Search[C]//IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16 - 22, 2011. AAAI Press, 2011.

[22] Zhen Y, Yeung D Y. A probabilistic model for multimodal hash function learning[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012: 940 - 948.

[23] Gong Y, Lazebnik S. Iterative quantization: A procrustean approach to learning binary codes[J]. IEEE Transactions on Multimedia, 2016, 16(2): 427 - 439.

[24] Tang J, Wang K, Shao L. Supervised matrix factorization hashing for cross-modal retrieval[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3157 - 3166.

(上接第 216 页)

[14] Eppstein D. Finding the k shortest paths[C]//Proceedings of the 35th Annual Symposium on Foundations of Computer Science. IEEE Computer Society, 1994: 154 - 165.

[15] Akiba T, Hayashi T, Nori N, et al. Efficient top-k shortest-path distance queries on large networks by pruned landmark labeling[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25 - 30, 2015, Austin, Texas, USA: AAAI Press, 2015: 349 - 360.

[16] Mislove A E. Online social networks: measurement, analysis, and applications to distributed information systems[D]. Rice University, 2009.

[17] SNAP: network datasets: Wikipedia vote network[EB/OL]. [2018 - 11 - 08]. http://snap.stanford.edu/data/wiki-Vote.html.