



(12) 发明专利

(10) 授权公告号 CN 109960732 B

(45) 授权公告日 2023.04.18

(21) 申请号 201910246991.4

G06N 3/0464 (2023.01)

(22) 申请日 2019.03.29

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 109960732 A

CN 103038750 A, 2013.04.10

CN 107092918 A, 2017.08.25

CN 107871014 A, 2018.04.03

(43) 申请公布日 2019.07.02

CN 109271486 A, 2019.01.25

(73) 专利权人 广东石油化工学院

US 2004003261 A1, 2004.01.01

地址 525000 广东省茂名市官渡二路139号  
大院

US 2015220684 A1, 2015.08.06

WO 2017092183 A1, 2017.06.08

(72) 发明人 荆晓远 董西伟 吴飞 黄鹤  
姚永芳 李云鹤

许胜等. 面向Web图像检索的基于语义迁移的无监督深度哈希.《计算机应用研究》.2018, 2866-2871.

(74) 专利代理机构 北京金智普华知识产权代理有限公司 11401

胡海峰;耿静静;冯巧遇;孙永;吴建盛. 哈希快速多标记学习算法. 信号处理. 2017, (08), 43-50.

专利代理师 杨采良

审查员 尤鑫

(51) Int. Cl.

G06F 16/432 (2019.01)

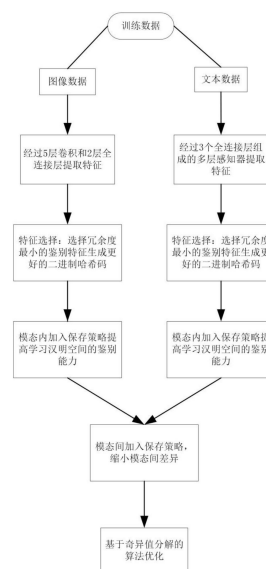
权利要求书3页 说明书15页 附图1页

(54) 发明名称

一种基于鲁棒监督的深度离散哈希跨模态检索方法及系统

(57) 摘要

本发明属于跨模态检索、模式识别技术领域,公开了一种基于鲁棒监督的深度离散哈希跨模态检索方法及系统,基于鲁棒监督的深度离散哈希跨模态检索方法采用模态内和模态间一致性保持策略,同时进行非冗余特征选择;并选择具有最小冗余度的鲁棒性和鉴别性特征生成二进制哈希码;再基于奇异值分解进行离散哈希学习。本发明的RSDDH方法可以优于几种最先进的浅层跨模态哈希方法,且随着哈希码长度的增加,本发明所提出的方法的性能就越好。根据实验结果,验证了特征选择策略、离散哈希方案,模态间和模态内一致性保持策略的有效性,提高了跨模态哈希检索的性能。



1. 一种基于鲁棒监督的深度离散哈希跨模态检索方法,其特点在于,所述基于鲁棒监督的深度离散哈希跨模态检索方法包括:对输入的图片或对应图片描述文本检索时,采用模态内和模态间一致性保持策略,同时进行非冗余特征选择;并选择具有最小冗余度的鲁棒性和鉴别性特征生成二进制哈希码;

再基于奇异值分解进行离散哈希学习;搜索出对应的图片或对应图片描述的文本;

所述基于鲁棒监督的深度离散哈希跨模态检索方法具体包括以下步骤:

步骤一:获取训练特征数据集,其中每个样本都由文本和图像对组成;

步骤二:对数据集里图像和文本分别进行特征学习;

步骤三:基于非冗余特征选择的深度哈希码学习;

步骤四:模态内加入一致性保持策略;

步骤五:模态间加入一致性保持策略;

步骤六:得到总目标函数并进行优化;

步骤二所述的对数据集里图像和文本分别进行特征学习的具体方法包括:

图像特征学习:利用一个7层的神经网络,前5层是卷积神经网络CNN,分别是第一层用96个大小为11\*11的卷积核,步长是4,0边距,用Relu激活函数;池化用传统的2\*2大小的模板做max pooling,最后局部归一化;第二层以第一层的输出作为输入,用256个5\*5大小的卷积核,步长是1,padding是2,同样用Relu激活函数,2\*2大小的max pooling,最后局部归一化;第三层、第四层和第五层相互连接,没有中间的池层或归一化层;第三和四层,用384个3\*3的卷积核,步长是1,padding是1,第五层用256个3\*3的卷积核,步长是1,padding是1;最大池层跟随第五层卷积的输出,得到尺寸为 $6 \times 6 \times 256$ 的输出特征;第6层全连接层有4096个神经元,激活函数用Relu;得到与输入图像对应的尺寸为 $4096 \times 1$ 的特征;最后连接第7层全连接层,激活函数用双曲正切TanH;

文本特征学习:采用多层感知器MLP,由三个全连接层构成一个MLP DNN,用于将文本特征从原始特征空间映射到语义空间;利用relu作为前两个完全连通层的非线性激活函数;并以双曲正切Tanh作为最后一层的激活函数;文本的维数是 $d^{(t)}$ ;

在步骤三中,基于非冗余特征选择的深度哈希码学习包括:对于第 $i$ 个实例 $(v_i, t_i)$ , $f(v_i; \theta_v) \in \mathbb{R}^{d^{(v)} \times 1}$ 表示学习到的图像特征,对应于图像模态的输出,其中 $\theta_v$ 为网络参数;

$F \in \mathbb{R}^{d^{(v)} \times n}$ 表示图像模态中 $n$ 个训练实例的深度学习特征矩阵,其中第 $i$ 个列向量为 $f(v_i; \theta_v)$ ;

$g(t_i; \theta_t) \in \mathbb{R}^{d^{(t)} \times 1}$ 表示学习到的文本特征,对应于文本模式的输出,其中 $\theta_t$ 为网络参数;  
 $G \in \mathbb{R}^{d^{(t)} \times n}$ 表示文本模态中 $n$ 个训练实例的深度学习特征矩阵,其中第 $i$ 个列向量为 $g(t_i; \theta_t)$ ;

图像和文本模态中的第 $i$ 个训练实例的投影深度学习特征 $P_v^T f(v_i; \theta_v)$ 和 $P_t^T g(t_i; \theta_t)$ 分别在汉明空间中产生二值哈希码 $b_i^{(v)}$ 和 $b_i^{(t)}$ ;  $P_v \in \mathbb{R}^{d^{(v)} \times k}$ 和 $P_t \in \mathbb{R}^{d^{(t)} \times k}$ 将深度学习特征 $f(v_i; \theta_v)$ 和 $g(t_i; \theta_t)$ 映射到潜空间的线性投影矩阵, $(\cdot)^T$ 是矩阵转置;

步骤三中,基于非冗余特征选择方法基于谱回归和 $2, 1$ -范数正则化进行非冗余特征选择,对图像模态和文本模式进行特征选择,获得图像和文本模式中的实例的二进制哈希码:

$$\begin{aligned}
\min_{\substack{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, \\ P_v, P_t, \theta_v, \theta_t}} \mathcal{J}_1 = & \sum_{i=1}^n \left\| P_v^T f(v_i; \theta_v) - b_i^{(v)} \right\|_2^2 + \gamma_1 \left( \left\| F^T P_v - Z^{(v)} \right\|_F^2 + \|P_v\|_{2,1} \right) \\
& + \sum_{i=1}^n \left\| P_t^T g(t_i; \theta_t) - b_i^{(t)} \right\|_2^2 + \gamma_2 \left( \left\| G^T P_t - Z^{(t)} \right\|_F^2 + \|P_t\|_{2,1} \right) \\
& + \gamma_3 \left( \left\| P_v^T F \mathbf{1} \right\|_F^2 + \left\| P_t^T G \mathbf{1} \right\|_F^2 \right) \\
\text{s.t. } & b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}
\end{aligned}$$

其中  $\gamma_1, \gamma_2, \gamma_3$  是权衡参数,  $\mathbf{1}$  表示所有元素为1的向量,  $\|\cdot\|_F$  表示矩阵的F-范数,  $Z^{(v)}$  和  $Z^{(t)}$  分别是基于两个相似矩阵构造的两个矩阵;  $Z^{(v)} = \hat{U}^{(v)} (\hat{\Sigma}^{(v)})^{1/2}$ ,  $\hat{\Sigma}^{(v)}$  和  $\hat{U}^{(v)}$  是通过相似矩阵  $\hat{S}^{(v)}$  进行奇异值分解SVD得到, 其中  $\hat{S}^{(v)} = \hat{U}^{(v)} \hat{\Sigma}^{(v)} (\hat{U}^{(v)})^T$ ,  $Z^{(t)} = \hat{U}^{(t)} (\hat{\Sigma}^{(t)})^{1/2}$ ,  $\hat{\Sigma}^{(t)}$  和  $\hat{U}^{(t)}$  通过对相似矩阵  $\hat{S}^{(t)}$  进行奇异值分解(SVD)得到; 用一组向量表示  $Z^{(v)} = [z_1^{(v)}, z_2^{(v)}, \dots, z_n^{(v)}]^T$  和  $Z^{(t)} = [z_1^{(t)}, z_2^{(t)}, \dots, z_n^{(t)}]^T$ ;  $\left\| F^T P_v - Z^{(v)} \right\|_F^2 + \|P_v\|_{2,1}$  和  $\left\| G^T P_t - Z^{(t)} \right\|_F^2 + \|P_t\|_{2,1}$  分别从图像模式和文本模式中选择鉴别特征和非冗余特征, 生成二进制哈希码。

2. 如权利要求1所述的基于鲁棒监督的深度离散哈希跨模态检索方法, 其特征在于, 步骤四中, 模态内加入一致性保持策略具体包括: 在图像模态中, 两个特征数据点  $v_i$  和  $v_j$  的模态内相似性  $S_{ij}^{(v)}$  为

$$S_{ij}^{(v)} = \begin{cases} 1, & \text{if } v_i \in N_{k_i}(v_j) \text{ or } v_j \in N_{k_j}(v_i) \\ 0, & \text{otherwise} \end{cases};$$

其中  $N_{k_i}(v_i)$  表示特征数据点  $v_i$  的K近邻集合; 文本模态中两个特征数据点  $t_i$  和  $t_j$  的模态内相似度  $S_{ij}^{(t)}$  为:

$$S_{ij}^{(t)} = \begin{cases} 1, & \text{if } t_i \in N_{k_i}(t_j) \text{ or } t_j \in N_{k_j}(t_i) \\ 0, & \text{otherwise} \end{cases};$$

其中  $N_{k_i}(t_i)$  表示特征数据点  $t_i$  的K近邻的集合; 每个特征数据点与特征数据点的K近邻在汉明空间中, 最小化问题为:

$$\begin{aligned}
\min_{\substack{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n}} \mathcal{J}_2 = & \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(v)} \left\| b_i^{(v)} - b_j^{(v)} \right\|_2^2 + \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(t)} \left\| b_i^{(t)} - b_j^{(t)} \right\|_2^2 \\
\text{s.t. } & b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}
\end{aligned}$$

3. 如权利要求1所述的基于鲁棒监督的深度离散哈希跨模态检索方法, 其特征在于, 步骤五中, 从图像模态数据点  $v_i$  ( $i=1, 2, \dots, n$ ) 和文本模态数据点  $t_j$ , 其中  $j=1, 2, \dots, n$ , 定义语义相关矩阵:

$$C_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } t_j \text{ have the same semantics} \\ 0, & \text{otherwise} \end{cases};$$

$v_i$  与  $t_j$  共享至少一个标签,  $v_i$  和  $t_j$  具有相同的语义,  $v_i^T v_j > 0$ ; 汉明空间中保持图像模态和文本模态之间的模态一致性中, 最小化问题表述如下:

$$\min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_3 = \sum_{i=1}^n \sum_{j=1}^n C_{ij} \|b_i^{(v)} - b_j^{(t)}\|_2^2。$$

$$\text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

4. 如权利要求1所述的基于鲁棒监督的深度离散哈希跨模态检索方法,其特征在于,步骤六中,得到总的目标函数

$$\min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, P_v, P_t, \theta_v, \theta_t} \mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3;$$

$$\text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1};$$

得到总目标函数并进行优化后,最后目标函数为:

$$\min_{B, P_v, P_t, \theta_v, \theta_t} \tilde{\mathcal{J}} = \|F^T P_v - B\|_F^2 + \gamma_1 \left( \|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1} \right)$$

$$+ \|G^T P_t - B\|_F^2 + \gamma_2 \left( \|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1} \right)$$

$$+ \gamma_3 \left( \|P_v^T F \mathbf{1}\|_F^2 + \|P_t^T G \mathbf{1}\|_F^2 \right) + 2tr(B^T L B)$$

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

其中  $B = [b_1, b_2, \dots, b_n]^T \in \{-1, +1\}^{n \times k}$ 。

5. 一种实施权利要求1所述基于鲁棒监督的深度离散哈希跨模态检索方法的基于鲁棒监督的深度离散哈希跨模态检索控制系统。

## 一种基于鲁棒监督的深度离散哈希跨模态检索方法及系统

### 技术领域

[0001] 本发明属于跨模态检索、模式识别技术领域,尤其涉及一种基于鲁棒监督的深度离散哈希跨模态检索方法及系统。

### 背景技术

[0002] 目前,业内常用的现有技术有3种浅层跨模态哈希方法:CMFH、SCM和SMFH,以及3种深度跨模态哈希方法:CDQ、PRDH和DCMH。:集体矩阵因式分解哈希(CMFH)方法,通过从一个实例的不同模态用潜在因子模型建立集体矩阵来学习一种统一的哈希码;语义关联最大化(SCM)方法,将语义标签集成到大规模数据建模的哈希学习过程中,利用所有监督的信息进行线性时间复杂度的训练,避免显式计算相似矩阵;集体深度量化(CDQ)方法,尝试在端到端深度结构中引入量化用于跨模态检索;关系对引导深度哈希(PRDH)方法,分别从模态内视图和模态间视图集成不同类型的成对约束来促进哈希码的相似性学习。

[0003] 数据往往以不同的方式表示。例如,在Flickr和Facebook网站等社交网络中,用户可以通过图片和相关文本记录事件。多模态数据迅速增加,同时在过去的几年里,技术的快速发展。跨模态检索的目的是将一种模态的数据作为查询,检索另一种模式的相关数据。考虑到跨模态检索的效率,哈希方法由于存储成本低和检索速度快,近年来受到了最近邻研究界的越来越多的关注。哈希的目的是将原始空间中的数据点映射到二进制哈希码的汉明空间中,在该空间中保留原始空间中的相似性。

[0004] 在哈希技术的基础上,提出了许多跨模态检索任务的实现方法。但这些方法通常采用传统的手工特征表示来进行跨模态检索,缺点是特征提取过程独立于哈希码学习过程,这意味着手工构造的特征可能与哈希码学习过程不兼容。因此,这些现有的具有手工制作特征的方法在实际应用中可能无法获得令人满意的性能。例如,局部二进制模式,尺度不变特征变换和梯度直方图,通常用于描述图像特征,缺点有二,其一该特征对于非线性特征提取的效果并不理想;其二通用性不好,这些方法提取的特征在一个数据集上适用但是换一个数据集适用性很差。至于文字,文档主题生成模型、重复软最大化模型和词袋模型通常用于描述文本特征,有两个缺点:首先需要人工进行特征工程,成本很高;其次所生成的文本表示是高纬度高稀疏的,并且特征表达能力很弱。

[0005] 综上所述,现有技术存在的问题是:

[0006] (1) 在图像和文本之间的检索中,传统的手工特征对不同的跨模态检索任务缺乏适应性和通用性,可能在某些数据集上表现很好,在其他数据集上所取得的跨模态检索性能通常不能令人满意。导致检索性能不理想的另一个原因可能是这些方法大多是基于浅层架构,它不能彻底利用有用的信息来完成特定的跨模态检索任务。

[0007] (2) 特征冗余问题。在各模态中,不同类型特征的判别能力是不同的,并且这些特征中存在冗余,一些现有技术尽管使用深度特征与特定的跨模态检索任务兼容,但是并没有采用特征选择策略,这样产生的哈希码所得到的跨模态检索性能存在一定的缺陷。

[0008] (3) 离散哈希学习问题。由于哈希码学习问题本质上是一个离散的学习问题,无法

轻易解决。许多现有的散列方法通过对离散约束进行放松并将问题转化为连续学习问题来解决离散学习问题。通过这种方式,尽管可以容易地解决离散学习问题,但是放松过程可能不利地影响所学习的二进制哈希码的准确性。

[0009] 解决上述技术问题的意义:

[0010] 本发明提出一种基于鲁棒监督的深度离散哈希跨模态检索方法,该方法是一种端到端的深度跨模态哈希方法,该方法同时采用了模态内,和模态间一致性保持策略,同时还进行了非冗余特征选择,可以选择具有最小冗余度的鲁棒性和鉴别性特征来生成更好的二进制哈希码,最后基于奇异值分解,解决离散哈希学习的问题。

## 发明内容

[0011] 针对现有技术存在的问题,本发明提供了一种基于鲁棒监督的深度离散哈希跨模态检索方法及系统。

[0012] 本发明提出一种基于鲁棒监督的深度离散哈希跨模态检索方法,该方法是一种端到端的深度跨模态哈希方法,可以用在图像和文本,之间的检索领域,当输入一张图片检索时,可以搜索出对应的图片或是对应图片描述的文本。不需要通告硬件,需要通过搜索引擎就可以完成该方法。同时采用了模态内,和模态间一致性保持策略,同时还进行了非冗余特征选择,可以选择具有最小冗余度的鲁棒性和鉴别性特征来生成更好的二进制哈希码,最后基于奇异值分解,解决离散哈希学习的问题。

[0013] 本发明是这样实现的,一种基于鲁棒监督的深度离散哈希跨模态检索方法,其特点在于,包括以下步骤:

[0014] 步骤1:获取训练特征数据集,其中每个样本都由文本和图像对组成;

[0015] 步骤2:对数据集里图像和文本分别进行特征学习;

[0016] 步骤3:基于非冗余特征选择的深度哈希码学习;

[0017] 步骤4:模态内加入一致性保存策略;

[0018] 步骤5:模态间加入一致性保存策略;

[0019] 步骤6:得到总目标函数并进行算法优化。

[0020] 进一步,步骤(2)所述的对数据集里图像和文本分别进行特征学习的具体方法是:

[0021] 图像特征学习:利用一个7层的神经网络,其中前5层是卷积神经网络(CNN),分别是第一层用96个大小为 $11 \times 11$ 的卷积核,步长是4,0边距(padding size),然后用Relu激活函数。池化用传统的 $2 \times 2$ 大小的模板做max pooling,最后局部归一化。第二层以第一层的输出作为输入,用256个 $5 \times 5$ 大小的卷积核,步长是1,padding是2,同样用Relu激活函数, $2 \times 2$ 大小的max pooling,最后局部归一化。第三层、第四层和第五层相互连接,没有中间的池层或归一化层。第三和四层,用384个 $3 \times 3$ 的卷积核,步长是1,padding是1,第五层用256个 $3 \times 3$ 的卷积核,步长是1,padding是1。最大池层跟随第五层卷积的输出,得到了尺寸为 $6 \times 6 \times 256$ 的输出特征。第6层全连接层有4096个神经元,为了防止过拟合,本发明设置dropout为0.5。激活函数用Relu。然后得到与输入图像对应的尺寸为 $4096 \times 1$ 的特征。最后连接第7层全连接层,激活函数用双曲正切(TanH)。

[0022] 文本特征学习:由于文本特征通常比图像特征更有判断力,所以文本特征与语义之间的关系能够更容易地建立是一种本能。本发明采用多层感知器(MLP),由三个全连接层

构成一个MLP DNN,用于将文本特征从原始特征空间映射到语义空间。利用relu作为前两个完全连通层的非线性激活函数。并以双曲正切(Tanh)作为最后一层的激活函数。文本的维数是 $d^{(t)}$ ,换句话说,最后一层有 $d^{(t)}$ 神经元。

[0023] 进一步,在步骤3中,基于非冗余特征选择的深度哈希码学习,对于第 $i$ 个实例 $(v_i, t_i)$ ,让 $f(v_i; \theta_v) \in \mathbb{R}^{d^{(v) \times 1}}$ 表示学习到的图像特征,它对应于图像模态的输出,其中 $\theta_v$ 网络参数。 $F \in \mathbb{R}^{d^{(v) \times n}}$ 表示图像模态中 $n$ 个训练实例的深度学习特征矩阵,其中第 $i$ 个列向量为 $f(v_i; \theta_v)$ 。此外, $g(t_i; \theta_t) \in \mathbb{R}^{d^{(t) \times 1}}$ 表示学习到的文本特征,它对应于文本模式的输出,其中 $\theta_t$ 网络参数。 $G \in \mathbb{R}^{d^{(t) \times n}}$ 表示文本模态中 $n$ 个训练实例的深度学习特征矩阵,其中第 $i$ 个列向量为 $g(t_i; \theta_t)$ 。假设图像和文本模态中的第 $i$ 个训练实例的投影深度学习特征 $P_v^T f(v_i; \theta_v)$ 和 $P_t^T g(t_i; \theta_t)$ 分别在汉明空间中产生二值哈希码 $b_i^{(v)}$ 和 $b_i^{(t)}$ 。其中 $P_v \in \mathbb{R}^{d^{(v) \times k}}$ 和 $P_t \in \mathbb{R}^{d^{(t) \times k}}$ 是将深度学习特征 $f(v_i; \theta_v)$ 和 $g(t_i; \theta_t)$ 映射到潜空间的线性投影矩阵, $(\cdot)^T$ 是矩阵转置。一般来说,对于图像和文本模式中的特征,都存在冗余特征。本发明提出了一种基于谱回归和 $2, 1$ -范数正则化的非冗余特征选择策略,用于对图像模态和文本模式进行特征选择。然后,本发明提出了以下优化问题,以获得图像和文本模式中的实例的二进制哈希码:

$$\min_{\substack{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, \\ P_v, P_t, \theta_v, \theta_t}} \mathcal{J}_1 = \sum_{i=1}^n \|P_v^T f(v_i; \theta_v) - b_i^{(v)}\|_2^2 + \gamma_1 \left( \|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1} \right)$$

$$\begin{aligned} [0024] \quad & + \sum_{i=1}^n \|P_t^T g(t_i; \theta_t) - b_i^{(t)}\|_2^2 + \gamma_2 \left( \|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1} \right) \\ & + \gamma_3 \left( \|P_v^T F \mathbf{1}\|_F^2 + \|P_t^T G \mathbf{1}\|_F^2 \right) \end{aligned}$$

$$[0025] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0026] 其中 $\gamma_1, \gamma_2, \gamma_3$ 是权衡参数, $\mathbf{1}$ 表示所有元素为1的向量, $\|\cdot\|_F$ 表示矩阵的F-范数, $Z^{(v)}$ 和 $Z^{(t)}$ 分别是基于两个相似矩阵构造的两个矩阵。 $Z^{(v)} = \hat{U}^{(v)}(\hat{\Sigma}^{(v)})^{1/2}$ ,  $\hat{\Sigma}^{(v)}$ 和 $\hat{U}^{(v)}$ 是通过相似矩阵 $\hat{S}^{(v)}$ 进行奇异值分解(SVD)得到,如: $\hat{S}^{(v)} = \hat{U}^{(v)}\hat{\Sigma}^{(v)}(\hat{U}^{(v)})^T$ 。同样的, $Z^{(t)} = \hat{U}^{(t)}(\hat{\Sigma}^{(t)})^{1/2}$ ,  $\hat{\Sigma}^{(t)}$ 和 $\hat{U}^{(t)}$ 是通过相似矩阵 $\hat{S}^{(t)}$ 进行奇异值分解(SVD)得到。为了方便描述,本发明用一组向量表示 $Z^{(v)}$ 和 $Z^{(t)}$ ,如: $Z^{(v)} = [z_1^{(v)}, z_2^{(v)}, \dots, z_n^{(v)}]^T$ 和 $Z^{(t)} = [z_1^{(t)}, z_2^{(t)}, \dots, z_n^{(t)}]^T$ 。总的来说, $\|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1}$ 和 $\|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1}$ 分别从图像模式和文本模式中选择鉴别特征和非冗余特征,以生成更好的二进制哈希码。

[0027] 进一步,在步骤4中,模态内加入一致性保存策略:模态内相似性可以反映每个模态中特征数据点之间的邻域关系,在图像模态中,两个特征数据点 $v_i$ 和 $v_j$ 的模态内相似性 $S_{ij}^{(v)}$ 可以定义为

$$[0028] \quad S_{ij}^{(v)} = \begin{cases} 1, & \text{if } v_i \in N_{k_i}(v_j) \text{ or } v_j \in N_{k_i}(v_i) \\ 0, & \text{otherwise} \end{cases}$$

[0029] 其中 $N_{k_i}(v_i)$ 表示特征数据点 $v_i$ 的 $K$ 近邻集合。同样,文本模态中两个特征数据点 $t_i$ 和 $t_j$ 的模态内相似度 $S_{ij}^{(t)}$ 也可以定义如下:

$$[0030] \quad S_{ij}^{(t)} = \begin{cases} 1, & \text{if } t_i \in N_{k_i}(t_j) \text{ or } t_j \in N_{k_i}(t_i) \\ 0, & \text{otherwise} \end{cases}$$

[0031] 其中  $N_{k_i}(t_i)$  表示特征数据点  $t_i$  的  $k$  近邻的集合。为了确保每个特征数据点与其  $k$  近邻在汉明空间中仍然相似,本发明将最小化问题表述如下:

$$[0032] \quad \min_{\substack{\{b_i^{(v)}\}_{i=1}^n, \\ \{b_i^{(t)}\}_{i=1}^n}} \mathcal{J}_2 = \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(v)} \|b_i^{(v)} - b_j^{(v)}\|_2^2 + \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(t)} \|b_i^{(t)} - b_j^{(t)}\|_2^2$$

$$[0033] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1} \quad \circ$$

[0034] 进一步,在步骤5中,模态间加入一致性保存策略,通过使用标签信息,本发明可以从图像模态数据点  $v_i$  ( $i=1, 2, \dots, n$ ) 和文本模态数据点  $t_j$  ( $j=1, 2, \dots, n$ ), 定义语义相关矩阵:

$$[0035] \quad C_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } t_j \text{ have the same semantics} \\ 0, & \text{otherwise} \end{cases}$$

[0036] 如果  $v_i$  与  $t_j$  共享至少一个标签,则认为  $v_i$  和  $t_j$  具有相同的语义,如:  $y_i^T y_j > 0$  是满足的。为了在汉明空间中保持图像模态和文本模态之间的模态一致性,本发明将最小化问题表述如下:

$$[0037] \quad \min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_3 = \sum_{i=1}^n \sum_{j=1}^n C_{ij} \|b_i^{(v)} - b_j^{(t)}\|_2^2$$

$$[0038] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0039] 进一步,在步骤6中,根据权利要求的3,4,5中,得到了总的目标函数

$$[0040] \quad \min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, P_v, P_t, \theta_v, \theta_t} \mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3$$

$$[0041] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0042] 现有的工作表明,如果在不同的模态空间中描述的数据具有相同的语义,他们被期望有相同的共同的潜在空间。在本文中,本发明假设图像和文本模态中具有相同语义的实例最终在低维公共汉明空间中用相同的二进制哈希码表示,则:  $b_i^{(v)} = b_i^{(t)} = b_i$  ( $i=1, 2, \dots, n$ ), 那么目标函数可以写为:

$$[0043] \quad \min_{\{b_i\}_{i=1}^n, P_v, P_t, \theta_v, \theta_t} \tilde{\mathcal{J}} = \tilde{\mathcal{J}}_1 + \tilde{\mathcal{J}}_2 + \tilde{\mathcal{J}}_3, \text{ s.t. } b_i \in \{-1, +1\}^{k \times 1}$$

[0044] 通过简单的推导,代数计算,即最后目标函数:

$$[0045] \quad \min_{B, P_v, P_t, \theta_v, \theta_t} \tilde{\mathcal{J}} = \|F^T P_v - B\|_F^2 + \gamma_1 \left( \|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1} \right) \\ + \|G^T P_t - B\|_F^2 + \gamma_2 \left( \|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1} \right) \\ + \gamma_3 \left( \|P_v^T F \mathbf{1}\|_F^2 + \|P_t^T G \mathbf{1}\|_F^2 \right) + 2 \text{tr}(B^T L B)$$

$$[0046] \quad \text{s.t. } B \in \{-1, +1\}^{n \times k}$$

$$[0047] \quad \text{其中 } B = [b_1, b_2, \dots, b_n]^T \in \{-1, +1\}^{n \times k}$$

[0048] 进一步,据本发明所知,目标函数中所有的变量  $B, P_v, P_t, \theta_v$  和  $\theta_t$  都不是凸的,本发明采用交替优化策略求解未知变量。换句话说,本发明每次更新一个变量时,其他变量都是



固定的。由于存在离散约束,很难直接求解目标函数方程中的未知二进制哈希码矩阵B的解,这里本发明提出了基于奇异值分解的离散哈希算法,求解这个未知的离散变量B。

[0049] 本发明的另一目的在于提供一种实施所述基于鲁棒监督的深度离散哈希跨模态检索方法的基于鲁棒监督的深度离散哈希跨模态检索控制系统。

[0050] 本发明的另一目的在于提供一种实施所述基于鲁棒监督的深度离散哈希跨模态检索方法的基于鲁棒监督的深度离散哈希跨模态检索网络平台。

[0051] 综上所述,本发明的优点及积极效果为:

[0052] 本发明表3统计了Wiki数据集上跨模态检索的mAP值。

[0053] 表4统计了NUS-WIDE数据集上跨模态检索的mAp值。

[0054] 表5统计了MIRFlickr数据集上跨模态检索的mAp值。

[0055] 表3.Wiki数据集上的mAP值

任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0056] 图片搜索 文本	CMFH	0.1900	0.2059	0.2014	0.1853
	SCM	0.2393	0.2303	0.2332	0.2510
	SMFH	0.2276	0.2060	0.2129	0.2321
	PRDH	0.2437	0.2372	0.2464	0.2600
	CDQ	0.2398	0.2495	0.2576	0.2629
	DCMH	0.2364	0.2559	0.2660	0.2642
	<b>RSDDH</b>	<b>0.2556</b>	<b>0.2909</b>	<b>0.2819</b>	<b>0.2801</b>
文本搜索 图片	CMFH	0.5875	0.6135	0.6053	0.6124
	SCM	0.3492	0.4035	0.3945	0.4046
	SMFH	0.6136	0.6453	0.6531	0.6090
	PRDH	0.6206	0.6296	0.6374	0.6469
	CDQ	0.6406	0.6485	0.6675	0.6681
	DCMH	0.6321	0.6338	0.6510	0.6566
	<b>RSDDH</b>	<b>0.6346</b>	<b>0.6808</b>	<b>0.6959</b>	<b>0.6976</b>

[0057] 表4.NUS-WIDE数据集上mAP值

任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0058] 图片搜索 文本	CMFH	0.4962	0.5006	0.5325	0.5314
	SCM	0.5097	0.5248	0.5434	0.5670
	SMFH	0.5524	0.5655	0.5676	0.5744
	PRDH	0.6104	0.6200	0.6318	0.6377
	CDQ	0.6055	0.6156	0.6251	0.6413
	DCMH	0.6139	0.6315	0.6439	0.6436
	<b>RSDDH</b>	<b>0.6509</b>	<b>0.6992</b>	<b>0.7179</b>	<b>0.7201</b>
文本搜索 图片	CMFH	0.5027	0.5106	0.5311	0.5405
	SCM	0.5188	0.5321	0.5515	0.5749
	SMFH	0.5749	0.5859	0.5959	0.5901
	PRDH	0.6102	0.6156	0.6338	0.6481
	CDQ	0.6457	0.6697	0.6319	0.6475
	DCMH	0.6321	0.6518	0.6625	0.6663
	<b>RSDDH</b>	<b>0.6536</b>	<b>0.7264</b>	<b>0.7250</b>	<b>0.7223</b>

[0059] 表5.MIRFlickr数据集上mAP值

任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0060] 图片搜索	CMFH	0.5883	0.5858	0.5891	0.5899
	SCM	0.6129	0.6345	0.6436	0.6581
	SMFH	0.5672	0.5645	0.5691	0.5506
	PRDH	0.6383	0.6633	0.6831	0.6710

[0061]	文本	CDQ	0.6477	0.6676	0.7005	0.7052
		DCMH	0.6728	0.6965	0.7101	0.7220
		<b>RSDDH</b>	<b>0.6909</b>	<b>0.7367</b>	<b>0.7503</b>	<b>0.7518</b>
	文本搜索 图片	CMFH	0.5876	0.5868	0.5919	0.5906
		SCM	0.5936	0.6081	0.6242	0.6482
		SMFH	0.6576	0.6503	0.6600	0.6643
		PRDH	0.6588	0.6672	0.6656	0.7320
CDQ		0.6968	0.7212	0.7134	0.7194	
DCMH		0.6860	0.7038	0.7288	0.7482	
<b>RSDDH</b>		<b>0.7383</b>	<b>0.7436</b>	<b>0.7827</b>	<b>0.8067</b>	

[0062] 通过观察表3,4,5本发明可以看出,与对比方法相比较本发明所提出的一种基于鲁棒监督的深度离散哈希跨模态检索方法在不同哈希代码长度下都具有较好的mAP值。实验结果还表明,本发明的RSDDH方法可以优于几种最先进的浅层跨模态哈希方法,且随着哈希码长度的增加,本发明所提出的方法的性能就越好。根据实验结果,验证了特征选择策略、离散哈希方案,模态间和模态内一致性保持策略的有效性,提高了跨模态哈希检索的性能。

#### 附图说明

[0063] 图1是本发明实施例提供的基于鲁棒监督的深度离散哈希跨模态检索方法流程图。

#### 具体实施方式

[0064] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0065] 现有技术中,尽管这些手工特征被广泛使用,对不同的跨模态检索任务缺乏适应性和通用性,结果,所取得的跨模态检索性能通常不能令人满意。导致检索性能不理想的另一个原因可能是这些方法大多是基于浅层架构,它不能彻底利用有用的信息来完成特定的跨模态检索任务。此外,在每一种模态中,不同类型特征的鉴别能力是不同的,这些特征中可能存在冗余。对于现有的不采用特征选择策略的跨模态哈希检索方法,可能会产生次优二进制哈希码。特别是对于现有的深度跨模态哈希检索方法,虽然学习到的深度特征和特定的跨模态检索任务是高度兼容的,但是可能也不会选择更合适和非冗余的特性而产生二进制哈希码。

[0066] 为解决上述技术问题,下面结合具体方案对本发明作详细描述。

[0067] 本发明实施例提供一种基于鲁棒监督的深度离散哈希跨模态检索方法,具体流程

如图1所示。

[0068] 步骤(1):获取训练特征数据集,其中每个样本都由文本和图像对组成,这里本发明用的是三个广泛使用的基准多模态数据集,分别是Wiki,MIRFlickrand NUS-WIDE。

[0069] 步骤(2):对数据集里图像和文本分别进行特征学习,具体方法是:

[0070] 图像特征学习模型参数设置(如表1所示):利用一个7层的神经网络,其中前5层是卷积神经网络(CNN),分别是第一层用96个大小为11\*11的卷积核,步长是4,0边距(padding size),然后用Relu激活函数。池化用传统的2\*2大小的模板做max pooling,最后局部归一化。第二层以第一层的输出作为输入,用256个5\*5大小的卷积核,步长是1,padding是2,同样用Relu激活函数,2\*2大小的max pooling,最后局部归一化。第三层、第四层和第五层相互连接,没有中间的池层或归一化层。第三和四层,用384个3\*3的卷积核,步长是1,padding是1,第五层用256个3\*3的卷积核,步长是1,padding是1。最大池层跟随第五层卷积的输出,得到了尺寸为6×6×256的输出特征。第6层全连接层有4096个神经元,为了防止过拟合,本发明设置dropout为0.5。激活函数用Relu。然后得到与输入图像对应的尺寸为4096×1的特征。最后连接第7层全连接层,激活函数用双曲正切(TanH)。

[0071] 表1

[0072]

Name	Type	Input Size	Filter number /size /stride /pad	Activation Function
Conv1	Convolution	227×227	96 /11×11 /4 /0	ReLU
Conv2	Convolution	27×27	256 /5×5 /1 /2	ReLU
Conv3	Convolution	13×13	384 /3×3 /1 /1	ReLU
Conv4	Convolution	13×13	384 /3×3 /1 /1	ReLU
Conv5	Convolution	13×13	256 /3×3 /1 /1	ReLU
Fc6	Fully connected	6×6	N/A	ReLU
Fc7	Fully connected	4096×1	N/A	TanH

[0073] 文本特征学习模型参数设置(如表2所示):由于文本特征通常比图像特征更有判别力,所以文本特征与语义之间的关系能够更容易地建立联系。本发明采用多层感知器(MLP),由三个全连接层构成一个MLP DNN,用于将文本特征从原始特征空间映射到语义空间。利用relu作为前两个完全连通层的非线性激活函数。并以双曲正切(Tanh)作为最后一层的激活函数。文本的维数是 $d^{(t)}$ ,换句话说,最后一层有 $d^{(t)}$ 神经元。

[0074] 表2

Name	Type	Input Size	Number of Nodes	Activation Function
Fc1	Fully connected	$s \times 1$	4096	ReLU
Fc2	Fully connected	$4096 \times 1$	4096	ReLU
Fc3	Fully connected	$4096 \times 1$	$d^{(t)}$	TanH

[0076] 步骤(3):基于非冗余特征选择的深度哈希码学习,对于第*i*个实例 $(v_i, t_i)$ ,让 $f(v_i; \theta_v) \in \mathbb{R}^{d^{(v)} \times 1}$ 表示学习到的图像特征,它对应于图像模态的输出,其中 $\theta_v$ 网络参数。 $F \in \mathbb{R}^{d^{(v)} \times n}$ 表示图像模态中*n*个训练实例的深度学习特征矩阵,其中第*i*个列向量为 $f(v_i; \theta_v)$ 。此外, $g(t_i; \theta_t) \in \mathbb{R}^{d^{(t)} \times 1}$ 表示学习到的文本特征,它对应于文本模态的输出,其中 $\theta_t$ 网络参数。 $G \in \mathbb{R}^{d^{(t)} \times n}$ 表示文本模态中*n*个训练实例的深度学习特征矩阵,其中第*i*个列向量为 $g(t_i; \theta_t)$ 。假设图像和文本模态中的第*i*个训练实例的投影深度学习特征 $P_v^T f(v_i; \theta_v)$ 和 $P_t^T g(t_i; \theta_t)$ 分别在汉明空间中产生二值哈希码 $b_i^{(v)}$ 和 $b_i^{(t)}$ 。其中 $P_v \in \mathbb{R}^{d^{(v)} \times k}$ 和 $P_t \in \mathbb{R}^{d^{(t)} \times k}$ 是将深度学习特征 $f(v_i; \theta_v)$ 和 $g(t_i; \theta_t)$ 映射到潜空间的线性投影矩阵, $(\cdot)^T$ 是矩阵转置。一般来说,对于图像和文本模态中的特征,都存在冗余特征。本发明提出了一种基于谱回归和 $2, 1$ -范数正则化的非冗余特征选择策略,用于对图像模态和文本模态进行特征选择。然后,本发明提出了以下优化问题,以获得图像和文本模态中的实例的二进制哈希码:

$$\begin{aligned}
 \min_{\substack{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, \\ P_v, P_t, \theta_v, \theta_t}} \mathcal{J}_1 = & \sum_{i=1}^n \left\| P_v^T f(v_i; \theta_v) - b_i^{(v)} \right\|_2^2 + \gamma_1 \left( \left\| F^T P_v - Z^{(v)} \right\|_F^2 + \left\| P_v \right\|_{2,1} \right) \\
 & + \sum_{i=1}^n \left\| P_t^T g(t_i; \theta_t) - b_i^{(t)} \right\|_2^2 + \gamma_2 \left( \left\| G^T P_t - Z^{(t)} \right\|_F^2 + \left\| P_t \right\|_{2,1} \right) \\
 & + \gamma_3 \left( \left\| P_v^T F \mathbf{1} \right\|_F^2 + \left\| P_t^T G \mathbf{1} \right\|_F^2 \right)
 \end{aligned}$$

$$[0078] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0079] 其中 $\gamma_1, \gamma_2, \gamma_3$ 是权衡参数, $\mathbf{1}$ 表示所有元素为1的向量, $\|\cdot\|_F$ 表示矩阵的F-范数, $Z^{(v)}$ 和 $Z^{(t)}$ 分别是基于两个相似矩阵构造的两个矩阵。 $Z^{(v)} = \hat{U}^{(v)}(\hat{\Sigma}^{(v)})^{1/2}$ ,  $\hat{\Sigma}^{(v)}$ 和 $\hat{U}^{(v)}$ 是通过相似矩阵 $\hat{S}^{(v)}$ 进行奇异值分解(SVD)得到,如: $\hat{S}^{(v)} = \hat{U}^{(v)}\hat{\Sigma}^{(v)}(\hat{U}^{(v)})^T$ 。同样的, $Z^{(t)} = \hat{U}^{(t)}(\hat{\Sigma}^{(t)})^{1/2}$ ,  $\hat{\Sigma}^{(t)}$ 和 $\hat{U}^{(t)}$ 是通过相似矩阵 $\hat{S}^{(t)}$ 进行奇异值分解(SVD)得到。为了描述,本发明用一组向量

表示 $Z^{(v)}$ 和 $Z^{(t)}$ ,如: $Z^{(v)}=[z_1^{(v)}, z_2^{(v)}, \dots, z_n^{(v)}]^T$ 和 $Z^{(t)}=[z_1^{(t)}, z_2^{(t)}, \dots, z_n^{(t)}]^T$ 。总的来说,  $\|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1}$ 和 $\|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1}$ 分别从图像模态和文本模态中选择鉴别特征和非冗余特征,以生成更好的二进制哈希码。

[0080] 步骤(4):模态内加入一致性保存策略:模态内相似性可以反映每个模态中特征数据点之间的邻域关系,在图像模态中,两个特征数据点 $v_i$ 和 $v_j$ 的模态内相似性 $S_{ij}^{(v)}$ 可以定义为

$$[0081] \quad S_{ij}^{(v)} = \begin{cases} 1, & \text{if } v_i \in N_{k_i}(v_j) \text{ or } v_j \in N_{k_i}(v_i) \\ 0, & \text{otherwise} \end{cases}$$

[0082] 其中 $N_{k_i}(v_i)$ 表示特征数据点 $v_i$ 的K近邻集合。同样,文本模态中两个特征数据点 $t_i$ 和 $t_j$ 的模态内相似度 $S_{ij}^{(t)}$ 也可以定义如下:

$$[0083] \quad S_{ij}^{(t)} = \begin{cases} 1, & \text{if } t_i \in N_{k_i}(t_j) \text{ or } t_j \in N_{k_i}(t_i) \\ 0, & \text{otherwise} \end{cases}$$

[0084] 其中 $N_{k_i}(t_i)$ 表示特征数据点 $t_i$ 的K近邻的集合。为了确保每个特征数据点与其K近邻在汉明空间中仍然相似,本发明将最小化问题表述如下:

$$[0085] \quad \min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_2 = \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(v)} \|b_i^{(v)} - b_j^{(v)}\|_2^2 + \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(t)} \|b_i^{(t)} - b_j^{(t)}\|_2^2$$

$$[0086] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0087] 步骤(5):模态间加入一致性保存策略,通过使用标签信息,本发明可以从图像模态数据点 $v_i$  ( $i=1, 2, \dots, n$ )和文本模态数据点 $t_j$  ( $j=1, 2, \dots, n$ ),定义语义相关矩阵:

$$[0088] \quad C_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } t_j \text{ have the same semantics} \\ 0, & \text{otherwise} \end{cases}$$

[0089] 如果 $v_i$ 与 $t_j$ 共享至少一个标签,则考虑 $v_i$ 和 $t_j$ 具有相同的语义,如: $v_i^T v_j > 0$ 是满足的。为了在汉明空间中保持图像模态和文本模态之间的模态一致性,本发明将最小化问题表述如下:

$$[0090] \quad \min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_3 = \sum_{i=1}^n \sum_{j=1}^n C_{ij} \|b_i^{(v)} - b_j^{(t)}\|_2^2$$

$$[0091] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0092] 步骤(6):根据权利要求的3,4,5中,得到了总的目标函数

$$[0093] \quad \min_{\{b_i^{(v)}\}_{i=1}^n, \{b_i^{(t)}\}_{i=1}^n, P_v, P_t, \theta_v, \theta_t} \mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3$$

$$[0094] \quad \text{s.t. } b_i^{(v)} \in \{-1, +1\}^{k \times 1}, b_i^{(t)} \in \{-1, +1\}^{k \times 1}$$

[0095] 现有的工作表明,如果在不同的模态空间中描述的数据具有相同的语义,他们被期望有相同的共同的潜在空间。在本发明中,本发明假设图像和文本模态中具有相同语义的实例最终在低维公共汉明空间中用相同的二进制哈希码表示,则:

$b_i^{(v)} = b_i^{(t)} = b_i$  ( $i=1, 2, \dots, n$ ),那么目标函数可以写为:

[0096] 
$$\min_{\{b_i\}_{i=1}^k, P_v, P_t, \theta_v, \theta_t} \tilde{\mathcal{J}} = \tilde{\mathcal{J}}_1 + \tilde{\mathcal{J}}_2 + \tilde{\mathcal{J}}_3, \text{ s.t. } b_i \in \{-1, +1\}^{k \times 1}$$

[0097] 通过简单的推导,代数计算,即最后目标函数:

[0098] 
$$\begin{aligned} \min_{B, P_v, P_t, \theta_v, \theta_t} \tilde{\mathcal{J}} = & \|F^T P_v - B\|_F^2 + \gamma_1 \left( \|F^T P_v - Z^{(v)}\|_F^2 + \|P_v\|_{2,1} \right) \\ & + \|G^T P_t - B\|_F^2 + \gamma_2 \left( \|G^T P_t - Z^{(t)}\|_F^2 + \|P_t\|_{2,1} \right) \\ & + \gamma_3 \left( \|P_v^T F \mathbf{1}\|_F^2 + \|P_t^T G \mathbf{1}\|_F^2 \right) + 2tr(B^T L B) \end{aligned}$$

[0099] s. t.  $B \in \{-1, +1\}^{n \times k}$

[0100] 其中  $B = [b_1, b_2, \dots, b_n]^T \in \{-1, +1\}^{n \times k}$

[0101] 在本发明实施例中,目标函数中所有的变量 $B, P_v, P_t, \theta_v$ 和 $\theta_t$ 都不是凸的,本发明采用交替优化策略求解未知变量。换句话说,本发明每次更新一个变量时,其他变量都是固定的。由于存在离散约束,很难直接求解目标函数方程中未知二进制哈希码矩阵 $B$ 的解,这里本发明提出了基于奇异值分解的离散哈希方法,求解这个未知的离散变量 $B$ 。

[0102] 为验证本方法是否有很好的优越性,本发明将提出的RSDDH方法与最近出版的几种最先进的跨模态哈希方法进行了比较,包括3个浅层的跨模态哈希方法,CMFH,SCM,SMFH和3个深度的跨模态哈希方法CDQ,PRDH,DCMH。分别在Wiki,MIRFlickr和NUS-WIDE三个数据集上进行了实验。

[0103] 下面结合实验结果对本发明作进一步描述。

[0104] 本发明表3统计了Wiki数据集上跨模态检索的mAP值。

[0105] 表4统计了NUS-WIDE数据集上跨模态检索的mAp值。

[0106] 表5统计了MIRFlickr数据集上跨模态检索的mAp值。

[0107] 表3.Wiki数据集上的mAP值



任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0108] 图片搜索 文本	CMFH	0.1900	0.2059	0.2014	0.1853
	SCM	0.2393	0.2303	0.2332	0.2510
	SMFH	0.2276	0.2060	0.2129	0.2321
	PRDH	0.2437	0.2372	0.2464	0.2600
	CDQ	0.2398	0.2495	0.2576	0.2629
	DCMH	0.2364	0.2559	0.2660	0.2642
	<b>RSDDH</b>	<b>0.2556</b>	<b>0.2909</b>	<b>0.2819</b>	<b>0.2801</b>
文本搜索 图片	CMFH	0.5875	0.6135	0.6053	0.6124
	SCM	0.3492	0.4035	0.3945	0.4046
	SMFH	0.6136	0.6453	0.6531	0.6090
	PRDH	0.6206	0.6296	0.6374	0.6469
	CDQ	0.6406	0.6485	0.6675	0.6681
	DCMH	0.6321	0.6338	0.6510	0.6566
	<b>RSDDH</b>	<b>0.6346</b>	<b>0.6808</b>	<b>0.6959</b>	<b>0.6976</b>

[0109] 表4.NUS-WIDE数据集上mAP值

任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0110] 图片搜索 文本	CMFH	0.4962	0.5006	0.5325	0.5314
	SCM	0.5097	0.5248	0.5434	0.5670
	SMFH	0.5524	0.5655	0.5676	0.5744
	PRDH	0.6104	0.6200	0.6318	0.6377
	CDQ	0.6055	0.6156	0.6251	0.6413
	DCMH	0.6139	0.6315	0.6439	0.6436
	<b>RSDDH</b>	<b>0.6509</b>	<b>0.6992</b>	<b>0.7179</b>	<b>0.7201</b>
文本搜索 图片	CMFH	0.5027	0.5106	0.5311	0.5405
	SCM	0.5188	0.5321	0.5515	0.5749
	SMFH	0.5749	0.5859	0.5959	0.5901
	PRDH	0.6102	0.6156	0.6338	0.6481
	CDQ	0.6457	0.6697	0.6319	0.6475
	DCMH	0.6321	0.6518	0.6625	0.6663
	<b>RSDDH</b>	<b>0.6536</b>	<b>0.7264</b>	<b>0.7250</b>	<b>0.7223</b>

[0111] 表5.MIRFlickr数据集上mAP值

任务	方法	哈希代码长度			
		16 位	32 位	64 位	128 位
[0112] 图片搜索 文本	CMFH	0.5883	0.5858	0.5891	0.5899
	SCM	0.6129	0.6345	0.6436	0.6581
	SMFH	0.5672	0.5645	0.5691	0.5506
	PRDH	0.6383	0.6633	0.6831	0.6710
	CDQ	0.6477	0.6676	0.7005	0.7052
	DCMH	0.6728	0.6965	0.7101	0.7220

	<b>RSDDH</b>	<b>0.6909</b>	<b>0.7367</b>	<b>0.7503</b>	<b>0.7518</b>
[0113] 文本搜索 图片	CMFH	0.5876	0.5868	0.5919	0.5906
	SCM	0.5936	0.6081	0.6242	0.6482
	SMFH	0.6576	0.6503	0.6600	0.6643
	PRDH	0.6588	0.6672	0.6656	0.7320
	CDQ	0.6968	0.7212	0.7134	0.7194
	DCMH	0.6860	0.7038	0.7288	0.7482
	<b>RSDDH</b>	<b>0.7383</b>	<b>0.7436</b>	<b>0.7827</b>	<b>0.8067</b>

[0114] 通过观察表3,4,5本发明可以看出,与对比方法相比较本发明所提出的一种基于鲁棒监督的深度离散哈希跨模态检索方法在不同哈希代码长度下都具有较好的mAP值。实验结果还表明,本发明的RSDDH方法可以优于几种最先进的浅层跨模态哈希方法,且随着哈希码长度的增加,本发明所提出的方法的性能就越好。根据实验结果,验证了特征选择策略、离散哈希方案,模态间和模态内一致性保持策略的有效性,提高了跨模态哈希检索的性能。

[0115] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

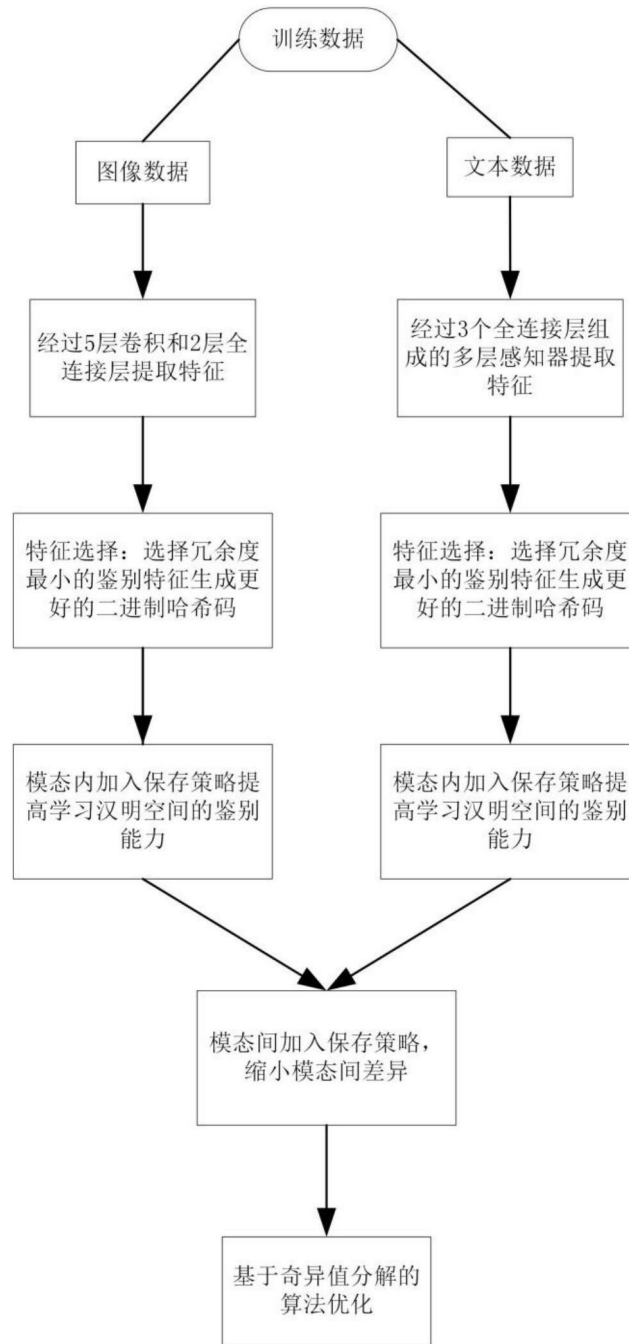


图1