(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110019652 B
(45)授权公告日 2022.06.03

(21)申请号 201910196009.7
(22)申请日 2019.03.14
(65)同一申请的已公布的文献号
申请公布号 CN 110019652 A
(43)申请公布日 2019.07.16
(73)专利权人 九江学院
地址 332000 江西省九江市前进东路551号
(72)发明人 董西伟 邓安远 周军 杨茂保 孙丽 胡芳 贾海英 王海霞
(74)专利代理机构 湖北创融蓝图知识产权代理事务所(特殊普通合伙) 42276
专利代理师 羊淑梅
(51)Int.Cl.
G06F 16/31(2019.01)
G06F 16/51(2019.01)

(56)对比文件
CN 109271486 A,2019.01.25
CN 105184303 A,2015.12.23
CN 109299342 A,2019.02.01
CN 109446347 A,2019.03.08
AU 1998085848 A1,1999.04.22
US 2017076143 A1,2017.03.16
CN 107885764 A,2018.04.06
CN 108170755 A,2018.06.15
CN 108334574 A,2018.07.27
董西伟.基于局部流形重构的半监督多视图图像分类.《计算机工程与应用》.2016,第52卷(第18期),

Yongming Chen 等.Continuum regression for cross-modal multimedia retrieval.《2012 19th IEEE International Conference on Image Processing》.2013,

Ran He 等.Cross-Modal Subspace Learning via Pairwise Constraints.《IEEE Transactions on Image Processing》.2015,第24卷

欧卫华 等.跨模态检索研究综述.《贵州师范大学学报(自然科学版)》.2018,第36卷(第02期),
(续)
审查员 陈曦

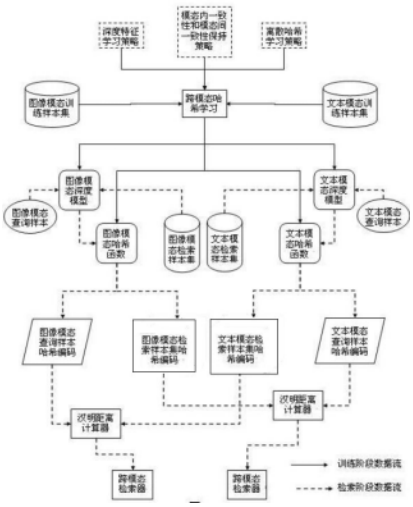
权利要求书2页 说明书13页 附图1页

(54)发明名称

一种基于深度学习的跨模态哈希检索方法

(57)摘要

一种基于深度学习的跨模态哈希检索方法，假设N个对象的图像模态的像素特征向量集为$V=\{v_i\in\Re^{m\times 1}\}_{i=1}^{N}$,其特征是该方法包括以下步骤：(1)使用基于深度学习技术设计的目标函数得到图像模态和文本模态共享的二进制哈希编码B，图像模态和文本模态的深度神经网络参数θ_v和θ_t,以及图像模态和文本模态的投影矩阵P_v和P_t；(2)使用交替更新的方式求解目标函数中的未知变量B、θ_v、θ_t、P_v和P_t；(3)基于求解得到的图像模态和文本模态的深度神经网络参数θ_v和θ_t,以及投影矩阵P_v和P_t；(4)基于生成的二进制哈希编码计算查询样本到检索样本集中各个样本的汉明距离；(5)使用基于近似最近邻搜索的跨模态检索器完成对查询样本的检索。该方法有效地提升了跨模态哈希检索的性能。



CN 110019652 B

[转续页]

[接上页]

(56)对比文件

姚伟娜.基于深度哈希算法的图像—文本跨模态检索研究.《中国优秀博硕士学位论文全文数据库(硕士)信息科技辑》.2019,(第01期),

1.一种基于深度学习的跨模态哈希检索方法，假设n个对象的图像模态的像素特征向量集为$V=\{v_i\in\Re^{r\times 1}\}_{i=1}^{n}$，其中，$v_i$表示第i个对象在图像模态的像素特征向量；令$T=\{t_i\in\Re^{s\times 1}\}_{i=1}^{n}$表示这n个对象在文本模态的特征向量，其中，$t_i$表示第i个对象在文本模态的特征向量；将n个对象的类别标记向量表示为$Y=\{y_i\in\Re^{c\times 1}\}_{i=1}^{n}$，其中，c表示对象类别的数量；对于向量$y_i$来说，如果第i个对象属于第k类，则令向量$y_i$的第k个元素为1，否则，向量$y_i$的第k个元素为0；其特征在于，该方法包括以下步骤：

(1)使用基于深度学习技术设计的目标函数得到图像模态和文本模态共享的二进制哈希编码B，图像模态和文本模态的深度神经网络参数θ_v和θ_t，以及图像模态和文本模态的投影矩阵P_v和P_t；

(2)使用交替求解的方式求解目标函数中的未知变量B、θ_v、θ_t、P_v和P_t，即交替的求解如下三个子问题：固定B、P_v和P_t，求解θ_v和θ_t；固定B、θ_v和θ_t，求解P_v和P_t；固定θ_v、θ_t、P_v和P_t，求解B；

(3)基于求解得到的图像模态和文本模态的深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t，为查询样本和检索样本集中的样本生成二进制哈希编码；

(4)基于生成的二进制哈希编码计算查询样本到检索样本集中各个样本的汉明距离；

(5)使用基于近似最近邻搜索的跨模态检索器完成对查询样本的检索；

所述步骤(1)中的基于深度学习技术设计的目标函数形式如下：

$$\min_{B,P_v,P_t,\theta_v,\theta_t}\tilde{J}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_1\left(\left\|P_v\right\|_F^2+\left\|P_t\right\|_F^2\right)$$
$$+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)+2tr(B^TLB),\ \text{s.t.}\ B\in\{-1,+1\}^{n\times k}\quad(1)$$

其中，γ_1和γ_2为非负平衡因子，$B=[b_1,b_2,\cdots,b_n]^T\in\{-1,+1\}^{n\times k}$，$P_v\in\Re^{d^{(v)}\times k}$和$P_t\in\Re^{d^{(t)}\times k}$为投影矩阵，$\theta_v$和$\theta_t$为深度神经网络参数，$F\in\Re^{d^{(v)}\times n}$和$G\in\Re^{d^{(t)}\times n}$分别为n个对象在图像模态和文本模态的深度特征，并且矩阵F和矩阵G的第i列的向量分别为$f(v_i;\theta_v)$和$g(t_i;\theta_t)$，$L\in\Re^{n\times n}$为拉普拉斯矩阵用于保持模态内一致性和模态间的一致性，1为全部元素为1的列向量，$||\cdot||_F$表示矩阵的Frobenius范数，tr(·)表示矩阵的迹，$(\cdot)^T$表示矩阵的转置。

2.根据权利要求1所述的一种基于深度学习的跨模态哈希检索方法，其特征在于，所述步骤(2)中的使用交替求解的方式求解目标函数中的未知变量B、θ_v、θ_t、P_v和P_t，具体为，交替地求解如下三个子问题：

(1)固定B、P_v和P_t，求解θ_v和θ_t；当固定二进制哈希编码B，以及投影矩阵P_v和P_t时，公式(1)所示的目标函数简化为关于深度神经网络参数θ_v和θ_t的子问题，即：

$$\min_{\theta_v,\theta_t}\tilde{J}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)\quad(2)；$$

(2)固定B、θ_v和θ_t，求解P_v和P_t；当固定二进制哈希编码B，以及深度神经网络参数θ_v和θ_t时，公式(1)所示的目标函数简化为关于投影矩阵P_v和P_t的子问题，即：

$$\min_{P_v,P_t}\tilde{J}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_1\left(\left\|P_v\right\|_F^2+\left\|P_t\right\|_F^2\right)+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)\quad(3)；$$

(3)固定θ_v、θ_t、P_v和P_t，求解B；当固定深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t时，

公式(1)所示的目标函数简化为关于二进制哈希编码B的子问题，即：

$$\min_{B} \tilde{J} = \left\| F^T P_v - B \right\|_F^2 + \left\| G^T P_t - B \right\|_F^2 + 2tr(B^T LB),\ \text{s.t.}\ B \in \{-1,+1\}^{n \times k} \quad (4)$$

在求解公式(4)中的未知变量B时，使用基于奇异值分解的离散哈希算法进行求解。

3.根据权利要求1所述的一种基于深度学习的跨模态哈希检索方法，其特征在于，所述步骤(3)中的基于求解得到的图像模态和文本模态的深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t，为查询样本和检索样本集中的样本生成二进制哈希编码，具体为，假设图像模态的一个查询样本的特征向量为$\tilde{v} \in \Re^{r \times 1}$，文本模态的一个查询样本的特征向量为$\tilde{t} \in \Re^{s \times 1}$，图像模态检索样本集中样本的特征为$\tilde{V} = \{\tilde{v}_i \in \Re^{r \times 1}\}_{i=1}^{\tilde{n}}$，文本模态检索样本集中样本的特征为$\tilde{T} = \{\tilde{t}_i \in \Re^{s \times 1}\}_{i=1}^{\tilde{n}}$，其中，$\tilde{n}$表示检索样本集中样本的数量；图像模态和文本模态查询样本和检索样本集中样本的二进制哈希编码分别为：$\tilde{b}^{(v)} = sign(P_v^T f(\tilde{v};\theta_v))$，$\tilde{b}^{(t)} = sign(P_v^T g(\tilde{t};\theta_t))$，$\tilde{b}_i^{(v)} = sign(P_v^T f(\tilde{v}_i;\theta_v))$和$\tilde{b}_i^{(t)} = sign(P_t^T g(\tilde{t}_i;\theta_t))$，其中，$i = 1,2,\cdots,\tilde{n}$，sign(•)为符号函数。

4.根据权利要求1所述的一种基于深度学习的跨模态哈希检索方法，其特征在于，所述步骤(4)中的基于生成的二进制哈希编码计算查询样本到检索样本集中各个样本的汉明距离，具体为，使用公式$d_{\tilde{v}\tilde{t}_i} = \tilde{b}^{(v)T}\tilde{b}_i^{(t)}$计算图像模态的查询样本到文本模态检索样本集中第i（$i = 1,2,\cdots,\tilde{n}$）个样本的汉明距离；使用公式$d_{\tilde{t}\tilde{v}_i} = \tilde{b}^{(t)T}\tilde{b}_i^{(v)}$计算文本模态的查询样本到图像模态检索样本集中第i（$i = 1,2,\cdots,\tilde{n}$）个样本的汉明距离。

5.根据权利要求1所述的一种基于深度学习的跨模态哈希检索方法，其特征在于，所述步骤(5)中的使用基于近似最近邻搜索的跨模态检索器完成对查询样本的检索，具体为，对计算得到的汉明距离$\{d_{\tilde{v}\tilde{t}_i}\}_{i=1}^{\tilde{n}}$或者$\{d_{\tilde{t}\tilde{v}_i}\}_{i=1}^{\tilde{n}}$按照从小到大的顺序进行排序，然后，在文本模态或者图像模态检索样本集中取前K个最小距离对应的样本作为检索结果。

一种基于深度学习的跨模态哈希检索方法

技术领域

[0001] 本发明涉及一种基于深度学习的跨模态哈希检索方法。

背景技术

[0002] 伴随着科学技术和社会生产力的快速发展,大数据时代悄然而至。所谓大数据是指在一定的时间范围内无法使用常规的软件工具进行捕捉、管理和处理的数据集合。IBM提出大数据具有5V特点,即:Volume(数据量大)、Variety(种类和来源多样化)、Value(数据价值密度相对较低,而有时却又弥足珍贵)、Velocity(数据增长速度快)、Veracity(数据的质量)。大数据也可以认为是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的信息资产。

[0003] 信息检索是数据处理的一个重要方面,而面对大数据,如何有效地进行信息检索成为大数据时代亟待解决并且非常具有挑战性的问题。对于大规模数据检索,哈希检索方法扮演着重要的角色。哈希检索方法将对象的高维特征映射到汉明空间中,生成一个低维的哈希编码来表示一个对象,它降低了检索系统对计算机内存空间的要求,提高了检索速度,能更好地适应海量检索的要求。哈希检索的主要思想是把高维向量表示的数据投影到汉明空间,在汉明空间中进行K近邻(K≥1)的检索。为了使汉明空间中的K近邻与原始空间保持一致,哈希学习算法需要满足局部保持特性,即,保持数据投影前后的相似性。局部敏感哈希(Locality Sensitive Hashing,LSH)方法可以使高维空间中距离很近的两点,在经过哈希函数对这两点进行哈希编码后,它们的哈希编码有很大的概率是一样的,反之,若两点之间的距离较远,则它们的哈希编码相同的概率会很小。

[0004] 跨模态哈希检索主要用于解决不同模态数据之间的相互检索问题,例如,用图像检索文本、或者用文本检索图像等。跨模态哈希检索方法需要对不同模态的数据进行哈希编码,生成紧凑的二进制哈希编码,然后基于生成的哈希编码完成不同模态数据之间的相互检索。Ding等人提出了集体矩阵分解哈希(Collective Matrix FactorizationHashing,CMFH)方法。CMFH方法可以利用集体矩阵分解从每个实例的不同模态学习统一的哈希编码。为了在基于矩阵分解的跨模态哈希方法中有效地使用类别信息并保持局部几何结构,进而达到有效提升由矩阵分解得到的潜在语义特征鉴别能力的目的,Tang等人提出了有监督矩阵分解哈希(Supervised Matrix Factorization Hashing,SMFH)方法。SMFH方法在进行哈希编码学习时,不仅考虑模态之间标记信息的一致性,还考虑模态内部局部几何结构的一致性。针对不少有监督跨模态哈希方法训练时间复杂度过高的问题,Zhang等人提出了称为语义相关性最大化(Semantic Correlation Maximization,SCM)的有监督跨模态哈希方法。SCM方法可以将语义标记信息无缝地集成到哈希学习过程中。

[0005] 上述的基于浅层学习结构的跨模态哈希算法所使用的手工特征可能不能与哈希编码学习到达最优的兼容性。为了解决这个问题,Jiang等人提出了深度跨模态哈希(Deep Cross-modal Hashing,DCMH)方法。DCMH方法是一种基于深度学习架构的端到端的跨模态哈希方法,它能够将特征学习和哈希编码学习有效地集成在一个学习框架中。为了在端到

端的学习架构中提升深度特征表示的量化能力(Quantizability),使得深度特征表示可以被更有效地量化,Cao等人通过将量化引入到用于跨模态检索的端到端深度学习架构中,提出了集体深度量化(Collective Deep Quantization,CDQ)方法。CDQ方法通过精心设计的混合网络和损失函数为两个模态联合学习深度特征表示和量化器。CDQ方法的混合网络包含:一个由多个卷积-池化(Convolution-Pooling)层构成的用于提取图像特征表示的图像网络,一个由多个全连接(Fully-Connected)层构成的用于提取文本特征表示的文本网络,两个用于生成最优低维特征表示的全连接瓶颈(Fully-Connected Bottleneck)层,一个用于捕获跨模态相关性的自适应交叉熵损失,以及一个用于控制哈希质量和量化能力的集体量化损失。此外,CDQ方法还可以学习模态共用的量化器码本,通过该码本可以实质性地增强两个模态之间的关联性。为了在用于跨模态检索的端到端的深度学习架构中有效地捕获不同模态之间的本质关系,Yang等人提出了成对关系导向的深度哈希(Pairwise RelationshipGuided Deep Hashing,PRDH)方法。PRDH方法从模态内的角度和模态间的角度通过集成不同类型的成对约束来学习更能反映模态间本质关系的哈希编码。此外,PRDH方法通过在深度学习架构中引入去相关约束来增强哈希编码每一比特的鉴别能力。

[0006] 跨模态哈希检索需要将对象在不同模态的高维特征数据映射到低维汉明空间,以实现基于汉明空间的二进制哈希编码快速、准确地完成跨模态信息检索任务。现有的跨模态哈希检索方法大多数是基于浅层学习结构的方法,这些方法虽然可以基于哈希检索技术快速地完成检索任务,但是浅层的学习结构使得原始特征中的鉴别信息不能够很好地被挖掘。深度学习技术在诸如分类任务、物体检测任务中已经展现出了优异的特征学习能力,并且现有的基于深度学习技术的跨模态哈希检索方法也表明深度学习技术对于提升跨模态检索任务的性能是有益的。因此,设计基于深度学习技术的跨模态哈希检索方法,对于完成大数据情境下的跨模态检索任务具有重要的意义和价值。

发明内容

[0007] 本发明其目的就在于提供一种基于深度学习的跨模态哈希检索方法,解决了现有的基于浅层学习结构的跨模态哈希检索方法不能够很好地挖掘原始特征中的鉴别信息的问题。

[0008] 为实现上述目的而采取的技术方案是,一种基于深度学习的跨模态哈希检索方法,假设n个对象的图像模态的像素特征向量集为$V=\{v_i\in\Re^{r\times 1}\}_{i=1}^{n}$,其中,$v_i$表示第i个对象在图像模态的像素特征向量;令$T=\{t_i\in\Re^{s\times 1}\}_{i=1}^{n}$表示这n个对象在文本模态的特征向量,其中,$t_i$表示第i个对象在文本模态的特征向量;将n个对象的类别标记向量表示为$Y=\{y_i\in\Re^{c\times 1}\}_{i=1}^{n}$,其中,c表示对象类别的数量;对于向量$y_i$来说,如果第i个对象属于第k类,则令向量$y_i$的第k个元素为1,否则,向量$y_i$的第k个元素为0;该方法包括以下步骤:

[0009] (1)使用基于深度学习技术设计的目标函数得到图像模态和文本模态共享的二进制哈希编码B,图像模态和文本模态的深度神经网络参数θ_v和θ_t,以及图像模态和文本模态的投影矩阵P_v和P_t;

[0010] (2)使用交替更新的方式求解目标函数中的未知变量B、θ_v、θ_t、P_v和P_t,即交替的求解如下三个子问题:固定B、P_v和P_t,求解θ_v和θ_t;固定B、θ_v和θ_t,求解P_v和P_t;固定θ_v、θ_t、P_v和

P_t，求解B；

[0011] (3)基于求解得到的图像模态和文本模态的深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t，为查询样本和检索样本集中的样本生成二进制哈希编码；

[0012] (4)基于生成的二进制哈希编码计算查询样本到检索样本集中各个样本的汉明距离；

[0013] (5)使用基于近似最近邻搜索的跨模态检索器完成对查询样本的检索。

[0014] 其中，所述步骤(1)中的基于深度学习技术设计的目标函数形式如下：

[0015] $$\begin{aligned}\min_{B,P_v,P_t,\theta_v,\theta_t}\tilde{\mathcal{J}} &= \left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_1\left(\left\|P_v\right\|_F^2+\left\|P_t\right\|_F^2\right)\\&+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)+2tr(B^TLB),\ \text{s.t.}\ B\in\{-1,+1\}^{n\times k}\end{aligned},\quad(1)$$

[0016] 其中，γ_1和γ_2为非负平衡因子，$B=[b_1,b_2,\cdots,b_n]^T\in\{-1,+1\}^{n\times k}$，$P_v\in\Re^{d^{(v)}\times k}$和$P_t\in\Re^{d^{(t)}\times k}$为投影矩阵，$\theta_v$和$\theta_t$为深度神经网络参数，$F\in\Re^{d^{(v)}\times n}$和$G\in\Re^{d^{(t)}\times n}$分别为n个对象在图像模态和文本模态的深度特征，并且矩阵F和矩阵G的第i列的向量分别为$f(v_i;\theta_v)$和$g(t_i;\theta_t)$，$L\in\Re^{n\times n}$为拉普拉斯矩阵用于保持模态内一致性和模态间的一致性，1为全部元素为1的列向量，$\|\cdot\|_F$表示矩阵的Frobenius范数，$tr(\cdot)$表示矩阵的迹，$(\cdot)^T$表示矩阵的转置。

[0017] 其中，所述步骤(2)中的使用交替更新的方式求解目标函数中的未知变量B、θ_v、θ_t、P_v和P_t，具体为，交替地求解如下三个子问题：

[0018] (1)固定B、P_v和P_t，求解θ_v和θ_t；当固定二进制哈希编码B，以及投影矩阵P_v和P_t时，公式(1)所示的目标函数简化为关于深度神经网络参数θ_v和θ_t的子问题，即：

[0019] $$\min_{\theta_v,\theta_t}\tilde{\mathcal{J}}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)\quad(2);$$

[0020] (2)固定B、θ_v和θ_t，求解P_v和P_t；当固定二进制哈希编码B，以及深度神经网络参数θ_v和θ_t时，公式(1)所示的目标函数简化为关于投影矩阵P_v和P_t的子问题，即：

[0021] $$\min_{P_v,P_t}\tilde{\mathcal{J}}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_1\left(\left\|P_v\right\|_F^2+\left\|P_t\right\|_F^2\right)+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)\quad(3);$$

[0022] (3)固定θ_v、θ_t、P_v和P_t，求解B；当固定深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t时，公式(1)所示的目标函数简化为关于二进制哈希编码B的子问题，即：

[0023] $$\min_{B}\tilde{\mathcal{J}}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+2tr(B^TLB),\ \text{s.t.}\ B\in\{-1,+1\}^{n\times k}\quad(4)$$

[0024] 在求解公式(4)中的未知变量B时，使用基于奇异值分解的离散哈希算法进行求解。

[0025] 其中，所述步骤(3)中的基于求解得到的图像模态和文本模态的深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t，为查询样本和检索样本集中的样本生成二进制哈希编码，具体为，假设图像模态的一个查询样本的特征向量为$\tilde{v}\in\Re^{r\times1}$，文本模态的一个查询样本的特征向量为$\tilde{t}\in\Re^{s\times1}$，图像模态检索样本集中样本的特征为$\tilde{V}=\{\tilde{v}_i\in\Re^{r\times1}\}_{i=1}^{\tilde{n}}$，文本模态检索样本集中样本的特征为$\tilde{T}=\{\tilde{t}_i\in\Re^{s\times1}\}_{i=1}^{\tilde{n}}$，其中，$\tilde{n}$表示检索样本集中样本的数量；图像模态和

文本模态查询样本和检索样本集中样本的二进制哈希编码分别为：$\tilde{b}^{(v)} = sign(P_v^T f(\tilde{v};\theta_v))$，$\tilde{b}^{(t)} = sign(P_v^T g(\tilde{t};\theta_t))$，$\tilde{b}_i^{(v)} = sign(P_v^T f(\tilde{v}_i;\theta_v))$和$\tilde{b}_i^{(t)} = sign(P_t^T f(\tilde{t}_i;\theta_t))$，其中，$i = 1,2,\cdots,\tilde{n}$，sign(•)为符号函数。

[0026] 其中，所述步骤(4)中的基于生成的二进制哈希编码计算查询样本到检索样本集中各个样本的汉明距离，具体为，使用公式$d_{\tilde{v}\tilde{t}_i} = \tilde{b}^{(v)T}\tilde{b}_i^{(t)}$计算图像模态的查询样本到文本模态检索样本集中第$i$（$i = 1,2,\cdots,\tilde{n}$）个样本的汉明距离；使用公式$d_{\tilde{t}\tilde{v}_i} = \tilde{b}^{(t)T}\tilde{b}_i^{(v)}$计算文本模态的查询样本到图像模态检索样本集中第$i$（$i = 1,2,\cdots,\tilde{n}$）个样本的汉明距离。

[0027] 其中，所述步骤(5)中的使用基于近似最近邻搜索的跨模态检索器完成对查询样本的检索，具体为，对计算得到的汉明距离$\{d_{\tilde{v}\tilde{t}_i}\}_{i=1}^{\tilde{n}}$（或者$\{d_{\tilde{t}\tilde{v}_i}\}_{i=1}^{\tilde{n}}$）按照从小到大的顺序进行排序，然后，在文本模态(或者图像模态)检索样本集中取前K个最小距离对应的样本作为检索结果。

[0028] 有益效果

[0029] 与现有技术相比本发明具有以下优点。

[0030] 1.本发明方法能够在保持检索速度的情况下，利用深度学习结构挖掘出更多的鉴别信息，从而可以更精准地完成跨模态检索；

[0031] 2.本发明方法通过实施模态内一致性和模态间一致性保持策略，充分地将原始特征空间的有益信息保持到汉明空间，促进了鉴别信息的挖掘和检索性能的提升；

[0032] 3.本发明方法所提出的基于奇异值分解的离散哈希算法可以使获取到的二进制哈希编码具有更多的有益特性，进而有效地提升了跨模态哈希检索的性能。

附图说明

[0033] 图1为本发明提出的基于深度学习的跨模态哈希检索方法工作流程图。

具体实施方式

[0034] 下面结合附图对本发明的技术方案做进一步的详细说明。

[0035] 本发明公开了一种基于深度学习的跨模态哈希检索方法，如图1所示，具体实施过程主要包括以下步骤：假设n个对象的图像模态的像素特征向量集为$V = \{v_i \in \Re^{r\times 1}\}_{i=1}^{n}$，其中，$v_i$表示第i个对象在图像模态的像素特征向量；令$T = \{t_i \in \Re^{s\times 1}\}_{i=1}^{n}$表示这n个对象在文本模态的特征向量，其中，$t_i$表示第i个对象在文本模态的特征向量；将n个对象的类别标记向量表示为$Y = \{y_i \in \Re^{c\times 1}\}_{i=1}^{n}$，其中，c表示对象类别的数量；对于向量$y_i$来说，如果第i个对象属于第k类，则令向量$y_i$的第k个元素为1，否则，向量$y_i$的第k个元素为0；

[0036] (1)基于深度学习的跨模态哈希检索目标函数构建

[0037] 本发明方法的目的是利用图像模态和文本模态的特征数据V和T，以及对象的类别标记信息学习图像模态和文本模态的哈希函数，并利用学习得到的哈希函数生成用于完成跨模态哈希检索任务的二进制哈希编码。直接使用图像模态和文本模态的特征数据V和T进

行跨模态哈希学习，不利于从原始特征中挖掘鉴别信息来生成性能优异的二进制哈希编码。为了更好地从图像模态和文本模态的原始特征数据中挖掘鉴别信息，本发明方法分别针对图像模态和文本模态数据构建深度神经网络(Deep Neural Network，DNN)进行深度特征学习。

[0038] 对于图像模态，本发明方法采用由AlexNet改进得到的由七个层构成的卷积神经网络(Convolutional Neural Network，CNN)进行图像模态深度特征学习。下面对这个CNN模型进行详细介绍。

[0039] 用于图像模态深度特征学习的这个CNN模型包含五个卷积层(Convolution Layer)和两个全连接层(Fully Connected Layer)，分别表示为“Conv1-Conv5”和“Fc6-Fc7”。该网络将图像模态的像素特征作为输入。在这个CNN中，第一个卷积层Conv1用96个大小为11×11×3核对大小为227×227×3的输入图像以4像素为步长进行过滤。在经过线性修正单元(Rectified Linear Unit，ReLU)的激活、最大池化(MAX-pooling)和局部响应归一化(Local Response Normalization，LRN)后，得到大小为27×27×96的输出特征。第二个卷积层Conv2以第一个卷积层Conv1的输出作为输入，Conv2用256个大小为5×5×96的核对输入进行过滤。同样地，在经过ReLU、MAX-pooling和LRN之后，得到大小为13×13×256的输出特征。第三、第四和第五个卷积层Conv3、Conv4和Conv5分别使用了大小为3×3×256、3×3×384和3×3×384的384、384和256个卷积核，并且每层都使用ReLU激活。当Conv5经过MAX-pooling后得到大小为6×6×256的输出特征。全连接层Fc6的神经元个数为4096，且使用0.5的丢弃比率对神经元进行暂时丢弃以防止过拟合。Fc7层为含有$d^{(v)}$个神经元的全连接层，并且用双曲正切(Hyperbolic Tangent，TanH)函数作为Fc7层的激活函数。最终，在Fc7层得到大小为$d^{(v)}\times 1$的输出特征。

[0040] 对于文本模态，本发明方法使用由三个全连接层构成的多层感知机(Multilayer Perceptron，MLP)来构建一个MLP深度神经网络将文本模态的特征从原始特征空间映射到语义空间。这里将所构建的MLP深度神经网络中的三个全连接层分别用Fc1、Fc2和Fc3表示。类似于相关文献进行文本模态特征学习时构建MLP深度神经网络的做法，本发明所构建的MLP深度神经网络的Fc1层和Fc2层使用ReLU作为非线性激活函数。对于Fc3层则使用双曲正切(TanH)函数作为激活函数。Fc3层的神经元个数为$d^{(t)}$，即：用于学习文本模态深度特征的MLP深度神经网络的输出特征的维数为$d^{(t)}$。

[0041] 对于第i个对象，令$f(v_i;\theta_v)\in\Re^{d^{(v)}\times 1}$表示图像模态的CNN的输出特征，其中，$\theta_v$为图像模态的CNN的参数；令$g(t_i;\theta_t)\in\Re^{d^{(t)}\times 1}$表示文本模态的MLP深度神经网络的输出特征，其中，θ_t为文本模态的MLP深度神经网络的参数。

[0042] 假设第i个对象在图像模态和文本模态的深度学习特征$f(v_i;\theta_v)$和$g(t_i;\theta_t)$经过线性投影矩阵$P_v\in\Re^{d^{(v)}\times k}$和$P_t\in\Re^{d^{(t)}\times k}$投影后的特征分别为$P_v^T f(v_i;\theta_v)$和$P_t^T g(t_i;\theta_t)$，其中，$(\cdot)^T$表示矩阵的转置。进一步假设由$P_v^T f(v_i;\theta_v)$和$P_t^T g(t_i;\theta_t)$可以分别生成汉明空间中的二进制哈希编码$b_i^{(v)}$和$b_i^{(t)}$。那么，可以通过如下的最小化问题进行跨模态哈希学习：

[0043] $$\min_{\substack{\{b_i^{(v)}\}_{i=1}^n,\{b_i^{(t)}\}_{i=1}^n,\\ P_v,P_t,\theta_v,\theta_t}} \mathcal{J}_1=\sum_{i=1}^{n}\left\|P_v^T f(v_i;\theta_v)-b_i^{(v)}\right\|_2^2+\sum_{i=1}^{n}\left\|P_t^T g(t_i;\theta_t)-b_i^{(t)}\right\|_2^2 +\gamma_1\left(\|P_v\|_F^2+\|P_t\|_F^2\right)+\gamma_2\left(\left\|P_v^T F\mathbf{1}\right\|_F^2+\left\|P_t^T G\mathbf{1}\right\|_F^2\right),\ \text{s.t.}\, b_i^{(v)}\in\{-1,+1\}^{k\times 1}, b_i^{(t)}\in\{-1,+1\}^{k\times 1} \quad (1)$$

[0044] 其中，γ_1和γ_2为非负平衡因子，等号右边的第三项为用于防止过拟合的正则项，等号右边第四项的作用是希望哈希编码的每一位是+1和-1的概率相等并用于最大化哈希编码的每一位所提供的信息。

[0045] 模态内相似性反映了每个模态内由特征向量构成的数据点之间的近邻关系。图像模态的两个数据点v_i和v_j之间的模态内相似性可以定义为：

[0046] $$S_{ij}^{(v)}=\begin{cases}\exp(-Edist_{ij}^{(v)}/2\sigma^2), \text{if } v_i\in N_{k_1}(v_j) \text{ or } v_j\in N_{k_1}(v_i)\\ 0, \text{otherwise}\end{cases}, \quad (2)$$

[0047] 其中，$N_{k_1}(v_i)$表示数据点v_i的k_1近邻集(k_1-nearest neighbors)，$Edist_{ij}^{(v)}$表示v_i和v_j之间的欧氏距离，即：$Edist_{ij}^{(v)}=\left\|v_i-v_j\right\|_2^2$。$\|\cdot\|_2$表示向量的$l_2$范数。σ用于控制$Edist_{ij}^{(v)}$的衰减速率。类似地，文本模态中由两个特征向量构成的两个数据点t_i和t_j的模态内相似性$S_{ij}^{(t)}$定义为：

[0048] $$S_{ij}^{(t)}=\begin{cases}\exp(-Edist_{ij}^{(t)}/2\sigma^2), \text{if } t_i\in N_{k_1}(t_j) \text{ or } t_j\in N_{k_1}(t_i)\\ 0, \text{otherwise}\end{cases}, \quad (3)$$

[0049] 其中，$Edist_{ij}^{(t)}=\left\|t_i-t_j\right\|_2^2$。对于每个模态，为了使数据点的局部近邻结构在汉明空间和原始特征空间保持一致，即：使原始特征空间中每个数据点及它近邻关系在汉明空间中得到保持，可以设计如下的目标函数：

[0050] $$\min_{\{b_i^{(v)}\}_{i=1}^n,\{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_2=\sum_{i=1}^{n}\sum_{j=1}^{n}S_{ij}^{(v)}\left\|b_i^{(v)}-b_j^{(v)}\right\|_2^2+\sum_{i=1}^{n}\sum_{j=1}^{n}S_{ij}^{(t)}\left\|b_i^{(t)}-b_j^{(t)}\right\|_2^2 \quad (4)$$
$$\text{s.t.}\, b_i^{(v)}\in\{-1,+1\}^{k\times 1}, b_i^{(t)}\in\{-1,+1\}^{k\times 1}$$

[0051] 基于对象的类别标记信息，可以定义图像模态的数据点v_i(i＝1,2,…,n)与文本模态的数据点t_j(j＝1,2,…,n)的如下所示的语义关联矩阵：

[0052] $$C_{ij}=\begin{cases}1, \text{if } v_i \text{ and } t_j \text{have the same semantics}\\ 0, \text{otherwise}\end{cases}。 \quad (5)$$

[0053] 需要说明的是：只要v_i和t_j至少属于一个相同的类别，则认为它们具有相同的语义。为了在汉明空间保持图像模态和文本模态之间的模态间一致性，可以设计如下的目标函数：

[0054] $$\min_{\{b_i^{(v)}\}_{i=1}^n,\{b_i^{(t)}\}_{i=1}^n} \mathcal{J}_3=\sum_{i=1}^{n}\sum_{j=1}^{n}C_{ij}\left\|b_i^{(v)}-b_j^{(t)}\right\|_2^2,\quad \text{s.t.}\, b_i^{(v)}\in\{-1,+1\}^{k\times 1}, b_i^{(t)}\in\{-1,+1\}^{k\times 1}。 \quad (6)$$

[0055] 综合以上关于图像模态深度特征学习、文本模态深度特征学习、模态内一致性和模态间一致性保持的分析，本发明方法的目标函数可以设计为：

[0056] $$\min_{\substack{\{b_i^{(v)}\}_{i=1}^n,\{b_i^{(t)}\}_{i=1}^n,\\ P_v,P_t,\theta_v,\theta_t}} \mathcal{J}=\mathcal{J}_1+\mathcal{J}_2+\mathcal{J}_3,\quad \text{s.t.}\, b_i^{(v)}\in\{-1,+1\}^{k\times 1}, b_i^{(t)}\in\{-1,+1\}^{k\times 1}。\ (7)$$

[0057] 根据已有工作，如果不同模态空间中的数据具有相同的语义，那么这些不同模态中的数据往往会对应一个公共的潜在空间。因此，本发明假设在图像模态和文本模态中具有相同语义的特征最终可以表示为公共汉明空间中相同的二进制哈希编码。也就是说有 $b_i^{(v)}=b_i^{(t)}=b_i$ （$i=1,2,\cdots,n$）成立。基于这个假设，公式(7)中的优化问题可以表示为：

[0058] $$\min_{\{b_i\}_{i=1}^n,P_v,P_t,\theta_v,\theta_t} \tilde{\mathcal{J}}=\tilde{\mathcal{J}}_1+\tilde{\mathcal{J}}_2+\tilde{\mathcal{J}}_3,\quad \text{s.t.}\, b_i\in\{-1,+1\}^{k\times 1},\qquad (8)$$

[0059] 其中，

[0060] $$\begin{aligned}\tilde{\mathcal{J}}_1=&\sum_{i=1}^n\left\|P_v^T f(v_i;\theta_v)-b_i\right\|_2^2+\sum_{i=1}^n\left\|P_t^T g(t_i;\theta_t)-b_i\right\|_2^2\\&+\gamma_1\left(\|P_v\|_F^2+\|P_t\|_F^2\right)+\gamma_2\left(\left\|P_v^T F\mathbf{1}\right\|_F^2+\left\|P_t^T G\mathbf{1}\right\|_F^2\right)\end{aligned},\qquad (9)$$

[0061] $$\tilde{\mathcal{J}}_2=\sum_{i=1}^n\sum_{j=1}^n S_{ij}^{(v)}\left\|b_i-b_j\right\|_2^2+\sum_{i=1}^n\sum_{j=1}^n S_{ij}^{(t)}\left\|b_i-b_j\right\|_2^2,\qquad (10)$$

[0062] $$\tilde{\mathcal{J}}_3=\sum_{i=1}^n\sum_{j=1}^n C_{ij}\left\|b_i-b_j\right\|_2^2。\qquad (11)$$

[0063] 通过简单的推导，$\tilde{\mathcal{J}}_1$ 可以重写为如下的形式：

[0064] $$\tilde{\mathcal{J}}_1=\left\|F^T P_v-B\right\|_F^2+\left\|G^T P_t-B\right\|_F^2+\gamma_1\left(\|P_v\|_F^2+\|P_t\|_F^2\right)+\gamma_2\left(\left\|P_v^T F\mathbf{1}\right\|_F^2+\left\|P_t^T G\mathbf{1}\right\|_F^2\right),\ (12)$$

[0065] 其中，$B=[b_1,b_2,\cdots,b_n]^T\in\{-1,+1\}^{n\times k}$，$F\in\Re^{d^{(v)}\times n}$，$G\in\Re^{d^{(t)}\times n}$，并且矩阵F和矩阵G的第i列的向量分别为$f(v_i;\theta_v)$和$g(t_i;\theta_t)$，$\|\cdot\|_F$表示矩阵的Frobenius范数。对 $\tilde{\mathcal{J}}_2+\tilde{\mathcal{J}}_3$ 进行如下的推导可以得到它的等价形式，即：

[0066] $$\begin{aligned}\tilde{\mathcal{J}}_2+\tilde{\mathcal{J}}_3=&\sum_{i=1}^n\sum_{j=1}^n S_{ij}^{(v)}\left\|b_i-b_j\right\|_2^2+\sum_{i=1}^n\sum_{j=1}^n S_{ij}^{(t)}\left\|b_i-b_j\right\|_2^2+\sum_{i=1}^n\sum_{j=1}^n C_{ij}\left\|b_i-b_j\right\|_2^2\\=&\sum_{i=1}^n\sum_{j=1}^n w_{ij}\left\|b_i-b_j\right\|_2^2=2tr(B^T LB)\end{aligned},\qquad (13)$$

[0067] 其中，$w_{ij}=S_{ij}^{(v)}+S_{ij}^{(t)}+C_{ij}$，L=D-W为拉普拉斯矩阵，$W\in\Re^{n\times n}$，$D\in\Re^{n\times n}$，$D_{ii}=\sum_{j=1}^n w_{ij}$ 表示对角矩阵D的第i个对角元素，w_{ij}为矩阵W第i行第j列上的元素，tr(•)表示矩阵的迹。根据公式(12)和公式(13)，公式(8)可以重写为：

[0068] $$\begin{aligned}\min_{B,P_v,P_t,\theta_v,\theta_t}\ &\tilde{\mathcal{J}}=\left\|F^T P_v-B\right\|_F^2+\left\|G^T P_t-B\right\|_F^2+\gamma_1\left(\|P_v\|_F^2+\|P_t\|_F^2\right)\\&+\gamma_2\left(\left\|P_v^T F\mathbf{1}\right\|_F^2+\left\|P_t^T G\mathbf{1}\right\|_F^2\right)+2tr(B^T LB),\ \text{s.t.}\, B\in\{-1,+1\}^{n\times k}\end{aligned}。\qquad (14)$$

[0069] (2)目标函数的求解

[0070] 公式(14)所示的目标函数中包含五个待求解的未知变量,即:二进制哈希编码矩阵B,线性投影矩阵P_v和P_t,深度神经网络参数θ_v和θ_t。公式(14)所示的目标函数对于这五个联合在一起的未知变量是非凸的,因此,无法同时得到这五个未知变量的解析解。公式(14)中的未知变量可以通过交替地求解如下三个子问题得到解,即:固定B、P_v和P_t,求解θ_v和θ_t;固定B、θ_v和θ_t,求解P_v和P_t;固定θ_v、θ_t、P_v和P_t,求解B。

[0071] (a)固定B、P_v和P_t,求解θ_v和θ_t

[0072] 当固定二进制哈希编码B,以及投影矩阵P_v和P_t时,公式(14)所示的目标函数简化为关于深度神经网络参数θ_v和θ_t的子问题,即:

[0073] $$\min_{\theta_v,\theta_t}\tilde{\mathcal{J}}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)。\qquad(15)$$

[0074] 本发明使用后向传播(Back Propagation,BP)算法来学习更新DNN网络参数θ_v。类似于大多数已有深度学习方法,这里使用基于后向传播的随机梯度下降算法来学习θ_v。学习θ_v的具体做法是:每次迭代从训练样本中选取一小批训练样本,然后利用选取的样本使用基于后向传播的随机梯度下降算法来学习θ_v。对于选取的训练样本的图像模态的每个特征向量v_i,首先使用如下的公式计算梯度:

[0075] $$\frac{\partial\tilde{\mathcal{J}}}{\partial f(v_i;\theta_v)}=2P_vP_v^Tf(v_i;\theta_v)+2\gamma_2P_vP_v^TF\mathbf{1}。\qquad(16)$$

[0076] 然后,使用链式规则和已得到的$\partial\tilde{\mathcal{J}}/\partial f(v_i;\theta_v)$计算$\partial\tilde{\mathcal{J}}/\partial\theta_v$。最后,利用计算得到的$\partial\tilde{\mathcal{J}}/\partial\theta_v$和BP算法更新图像模态的DNN网络参数$\theta_v$。

[0077] 算法1展示了求解图像模态DNN网络参数θ_v的算法。

算法 1：求解图像模态 DNN 网络参数θ_v的算法

输入：图像模态特征集V，投影矩阵P_v，二进制哈希编码矩阵B。

输出：DNN 网络参数θ_v。

步骤：

1.初始化 DNN 网络参数θ_v、每批特征向量的数量N_v和迭代次数$iter_v=\lceil n/N_v \rceil$；

[0078]

2.for $i=1,2,\cdots,iter_v$ do

3. 从图像模态特征集V中随机选择N_v个特征向量；

4. 对每个特征向量v_i，使用前向传播算法计算$f(v_i;\theta_v)$；

5. 使用公式（16）计算$\partial\tilde{J}/\partial f(v_i;\theta_v)$；

6. 使用后向传播更新 DNN 网络参数θ_v；

7.end for

[0079] 类似地，使用基于反向传播的随机梯度下降算法学习更新文本模态的深度神经网络参数θ_t。对于选取的训练样本的文本模态的每个特征向量t_i，首先计算如下的梯度：

[0080] $$\frac{\partial\tilde{J}}{\partial g(t_i;\theta_t)}=2P_tP_t^Tg(t_i;\theta_t)+2\gamma_2P_tP_t^TG\mathbf{1}。\qquad(17)$$

[0081] 然后，使用链式规则和已得到的梯度$\partial\tilde{J}/\partial g(t_i;\theta_t)$计算$\partial\tilde{J}/\partial\theta_t$。最后，利用计算得到的$\partial\tilde{J}/\partial\theta_t$和BP算法更新文本模态的DNN网络参数$\theta_t$。使用与算法1类似的算法可以学习得到文本模态的DNN网络参数θ_t。

[0082] (b)固定B、θ_v和θ_t，求解P_v和P_t

[0083] 当固定二进制哈希编码B，以及深度神经网络参数θ_v和θ_t时，公式(14)所示的目标函数简化为关于投影矩阵P_v和P_t的子问题，即：

[0084] $$\min_{P_v,P_t}\tilde{J}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+\gamma_1\left(\left\|P_v\right\|_F^2+\left\|P_t\right\|_F^2\right)+\gamma_2\left(\left\|P_v^TF\mathbf{1}\right\|_F^2+\left\|P_t^TG\mathbf{1}\right\|_F^2\right)。(18)$$

[0085] 针对公式(18)中的$\tilde{J}$分别关于P_v和P_t求偏导数并令偏导数等于0，可以得到：

[0086] $$\frac{\partial\tilde{J}}{\partial P_v}=2F(F^TP_v-B)+2P_v+2\gamma_2F11^TF^TP_v=0，\qquad(19)$$

[0087] $$\frac{\partial \tilde{\mathcal{J}}}{\partial P_t}=2G(G^TP_t-B)+2P_t+2\gamma_2G11^TG^TP_t=0\text{。}\qquad(20)$$

[0088] 经过简单的推导可得：

[0089] $P_v=(FF^T+I+F11^TF^T)^{-1}FB$， (21)

[0090] $P_t=(GG^T+I+G11^TG^T)^{-1}GB$， (22)

[0091] 其中，I为单位矩阵，$(\bullet)^{-1}$表示矩阵的逆。

[0092] (c)固定θ_v、θ_t、P_v和P_t，求解B

[0093] 当固定深度神经网络参数θ_v和θ_t，以及投影矩阵P_v和P_t时，公式(14)所示的目标函数简化为关于二进制哈希编码B的子问题，即：

[0094] $$\min_B \tilde{\mathcal{J}}=\left\|F^TP_v-B\right\|_F^2+\left\|G^TP_t-B\right\|_F^2+2tr(B^TLB),\ \text{s.t.}\ B\in\{-1,+1\}^{n\times k}\text{。}\qquad(23)$$

[0095] 对公式(23)进行简单的推导可以得到：

[0096] $$\min_B \tilde{\mathcal{J}}=\left\|F^TP_v\right\|_F^2-2tr(BP_v^TF)+\left\|B\right\|_F^2+\left\|G^TP_t\right\|_F^2-2tr(BP_t^TG)+\left\|B\right\|_F^2+2tr(B^TLB)\quad\text{s.t.}\ B\in\{-1,+1\}^{n\times k}\text{。}\qquad(24)$$

[0097] 因为P_v、P_t、θ_v和θ_t是固定的，因此，$\left\|F^TP_v\right\|_F^2$和$\left\|G^TP_t\right\|_F^2$均为常数。进一步，在公式(24)中忽略这两项也不会对B的求解产生影响。此外，因为$B\in\{-1,+1\}^{n\times k}$，可以得到$\left\|B\right\|_F^2=nk$，也就说，$\left\|B\right\|_F^2$为常数。将公式(24)中的常数项舍弃后，公式(24)转化为：

[0098] $$\min_B \tilde{\mathcal{J}}=tr(B^TLB)-tr(BQ),\quad\text{s.t.}\ B\in\{-1,+1\}^{n\times k}\text{，}\qquad(25)$$

[0099] 其中，$Q=P_v^TF+P_t^TG$。

[0100] 公式(25)中的未知变量为离散变量，因此，一般情况下很难直接对其进行求解得到解析解。本发明提出了基于奇异值分解的离散哈希算法来求解公式(25)所示的关于离散变量B的优化问题。下面详细介绍基于奇异值分解的离散哈希算法。

[0101] 对矩阵L进行奇异值分解可以得到$L=\tilde{U}\Sigma\tilde{V}^T$，其中，$\tilde{U}\in\Re^{n\times n}$，$\tilde{V}\in\Re^{n\times n}$，$\Sigma\in\Re^{n\times n}$为对角矩阵。将$L=\tilde{U}\Sigma\tilde{V}^T$代入公式(25)可以得到：

[0102] $$\min_B \tilde{\mathcal{J}}=tr(B^T\tilde{U}\Sigma\tilde{V}^TB)-tr(BQ),\quad\text{s.t.}\ B\in\{-1,+1\}^{n\times k}\text{。}\qquad(26)$$

[0103] 令$b_i^T\in\{-1,+1\}^{1\times k}$、$\tilde{u}_i^T\in\Re^{1\times n}$和$\tilde{v}_i^T\in\Re^{1\times n}$分别表示矩阵B、$\tilde{U}$和$\tilde{V}$的第i行；令$B_{-i}\in\Re^{(n-1)\times k}$、$\tilde{U}_{-i}\in\Re^{(n-1)\times n}$和$\tilde{V}_{-i}\in\Re^{(n-1)\times n}$分别表示矩阵B、$\tilde{U}$和$\tilde{V}$在去除了$b_i^T$、$\tilde{u}_i^T$和$\tilde{v}_i^T$之后剩余的行构成的矩阵。此时，可以得到：

[0104] $$\begin{aligned}tr(B^T\tilde{U}\Sigma\tilde{V}^TB)&=tr((\tilde{u}_ib_i^T+\tilde{U}_{-i}^TB_{-i})\Sigma(\tilde{v}_ib_i^T+\tilde{V}_{-i}^TB_{-i}))\\&=tr(b_i(\tilde{u}_i^T\Sigma\tilde{V}_{-i}^T+\tilde{v}_i^T\Sigma\tilde{U}_{-i}^T))+\text{const}\end{aligned}\text{。}\qquad(27)$$

[0105] 类似地，可以得到：

[0106] $$tr(BQ)=tr(QB)=tr(q_ib_i^T+Q_{-i}B_{-i})=tr(q_ib_i^T)+\text{const}\text{。}\qquad(28)$$

[0107] 其中，$q_i \in \Re^{k\times 1}$ 表示矩阵Q的第i列，$Q_{-i} \in \Re^{k\times(n-1)}$ 表示矩阵Q在去除q_i后剩余的列构成的矩阵。

[0108] 根据公式(27)和公式(28)，公式(26)中未知的二进制哈希编码矩阵B可以通过求解如下关于b_i (i＝1,2,…,n)的优化问题得到，即：

[0109] $$\min_{b_i} tr(b_i(\tilde{u}_i^T\Sigma\tilde{V}_{-i}^T+\tilde{v}_i^T\Sigma\tilde{U}_{-i}^T))-tr(q_ib_i^T),\quad \text{s.t.}\, b_i\in\{-1,+1\}^{k\times 1}。 \qquad (29)$$

[0110] 经过简单的推导，公式(29)可以转化为：

[0111] $$\min_{b_i} b_i^T(B_{-i}^T(\tilde{V}_{-i}\Sigma\tilde{u}_i+\tilde{U}_{-i}\Sigma\tilde{v}_i)-q_i),\quad \text{s.t.}\, b_i\in\{-1,+1\}^{k\times 1}。 \qquad (30)$$

[0112] 公式(30)中的优化问题具有如下的解析解：

[0113] $$b_i=sign(q_i-B_{-i}^T(\tilde{V}_{-i}\Sigma\tilde{u}_i+\tilde{U}_{-i}\Sigma\tilde{v}_i))。 \qquad (31)$$

[0114] 其中，sign(•)表示符号函数。

[0115] 算法2展示了基于奇异值分解的离散哈希算法。

[0116]

算法 2：基于奇异值分解的离散哈希算法

输入：图像模态深度特征矩阵 F，文本模态深度特征矩阵 G，投影矩阵 P_v 和 P_t，拉普拉斯矩阵 L。

输出：二进制哈希编码矩阵 $B=[b_1,b_2,\cdots,b_i,\cdots,b_n]^T$。

步骤：

1.计算 $Q=P_v^TF+P_t^TG$；

2.通过对矩阵 L 进行奇异值分解获取矩阵 $\tilde{U}$、$\tilde{V}$ 和 Σ；

3.for $i=1,2,\cdots,n$ do

4. 分别基于 $\tilde{U}$、$\tilde{V}$ 和 Q 获取 $\tilde{u}_i$、$\tilde{v}_i$ 和 q_i；

5. 分别基于 B、$\tilde{U}$ 和 $\tilde{V}$ 获取 B_{-i}、$\tilde{U}_{-i}$ 和 $\tilde{V}_{-i}$；

6. 使用公式（31）计算 b_i；

7.end for

[0117] (3)生成查询样本和检索样本集中的样本二进制哈希编码

[0118] 假设图像模态的一个查询样本的特征向量为 $\tilde{v}\in\Re^{r\times 1}$，文本模态的一个查询样本的特征向量为 $\tilde{t}\in\Re^{s\times 1}$，图像模态检索样本集中样本的特征为 $\tilde{V}=\{\tilde{v}_i\in\Re^{r\times 1}\}_{i=1}^{\tilde{n}}$，文本模态检索样本集中样本的特征为 $\tilde{T}=\{\tilde{t}_i\in\Re^{s\times 1}\}_{i=1}^{\tilde{n}}$，其中，$\tilde{n}$ 表示检索样本集中样本的数量。利用求

解得到的图像模态和文本模态的投影矩阵P_v和P_t，以及图像模态和文本模态的深度神经网络参数θ_v和θ_t，可以得到图像模态和文本模态查询样本和检索样本集中样本的二进制哈希编码分别为：$\tilde{b}^{(v)}=sign(P_v^T f(\tilde{v};\theta_v))$，$\tilde{b}^{(t)}=sign(P_v^T g(\tilde{t};\theta_t))$，$\tilde{b}_i^{(v)}=sign(P_v^T f(\tilde{v}_i;\theta_v))$和$\tilde{b}_i^{(t)}=sign(P_t^T f(\tilde{t}_i;\theta_t))$，其中，$i=1,2,\cdots,\tilde{n}$，sign(•)为符号函数。

[0119] (4)计算查询样本到检索样本集中各个样本的汉明距离

[0120] 对于图像模态的查询样本$\tilde{v}$，使用公式$d_{\tilde{v}\tilde{t}_i}=\tilde{b}^{(v)T}\tilde{b}_i^{(t)}$（$i=1,2,\cdots,\tilde{n}$）计算图像模态的查询样本$\tilde{v}$到文本模态检索样本集中样本$\tilde{t}_i$（$i=1,2,\cdots,\tilde{n}$）的汉明距离。对于文本模态的查询样本$\tilde{t}$，使用公式$d_{\tilde{t}\tilde{v}_i}=\tilde{b}^{(t)T}\tilde{b}_i^{(v)}$（$i=1,2,\cdots,\tilde{n}$）计算文本模态的查询样本$\tilde{t}$到图像模态检索样本集中样本$\tilde{v}_i$（$i=1,2,\cdots,\tilde{n}$）的汉明距离。

[0121] (5)使用跨模态检索器完成对查询样本的检索

[0122] 对于图像检索文本的检索任务，首先对计算得到的$\tilde{n}$个汉明距离$d_{\tilde{v}\tilde{t}_1},d_{\tilde{v}\tilde{t}_2},\cdots,d_{\tilde{v}\tilde{t}_{\tilde{n}}}$按照从小到大的顺序进行排序，然后，在文本检索样本集中取前K个最小距离对应的样本作为检索结果。类似地，对于文本检索图像的检索任务，首先对计算得到的$\tilde{n}$个汉明距离$d_{\tilde{t}\tilde{v}_1},d_{\tilde{t}\tilde{v}_2},\cdots,d_{\tilde{t}\tilde{v}_{\tilde{n}}}$按照从小到大的顺序进行排序，然后，在图像检索样本集中取前K个最小距离对应的样本作为检索结果。

[0123] 以下结合具体实验对本发明的有益效果进行说明。

[0124] 针对本发明方法所实施的相关实验主要在Pascal VOC 2007数据集上进行，首先简要介绍Pascal VOC 2007数据集。Pascal VOC 2007数据集包含属于20个类别(例如，飞机、瓶子、马和沙发等)的9963幅图像，并且每幅图像均被标记了标签。在实验中，本发明方法将数据集划分成包含5011个图像-标签对的训练集和包含4952个图像-标签对的测试集。对于深度跨模态哈希方法，图像模态使用原始像素特征作为输入特征。对于以手工特征作为输入的方法，使用512维的GIST特征作为输入特征。对于文本模态，使用399维的词频特征作为输入特征。在实验中主要进行两种跨模态检索任务，即：用图像检索文本和用文本检索图像，分别用Img2Txt和Txt2Img表示。

[0125] 本发明使用平均精度均值(Mean Average Precision，MAP)来衡量跨模态哈希检索方法的性能。为了获取MAP，需要首先针对每个查询样本计算平均精度(Average Precision，AP)。当获取了所有查询样本的平均精度AP后，对所有平均精度AP求平均值即可得到平均精度均值MAP。

[0126] 本发明方法使用动量(Momentum)和权值衰减(Weight Decay)分别为0.9和0.0001的小批量梯度下降算法，并且批(Batch)的大小设置为128。使用在ImageNet数据集上预训练的AlexNet来初始化本发明方法中图像模态深度神经网络的前五层。对于本发明方法中深度神经网络的其它参数采用随机初始化的方式进行初始化。将图像模态和文本模态的深度神经网络的输出特征维数均设置为1024。在实验中，采用5折交叉验证来确定本发明方法中参数γ_1和γ_2的最佳值。对于其它方法中的参数，按照各个方法所推荐的参数设置原则进行参数设置，实验所报告的结果为10次随机实验结果的平均值。

[0127] 与本发明方法进行对比的方法分别为：语义相关性最大化(Semantic

Correlation Maximization,SCM)、有监督矩阵分解哈希(Supervised Matrix Factorization Hashing,SMFH)方法、深度跨模态哈希(Deep Cross-Modal Hashing,DCMH)方法和成对关系导向的深度哈希(Pairwise Relationship Guided Deep Hashing,PRDH)方法。表1列出了本发明方法和对比方法在Pascal VOC 2007数据集上进行跨模态哈希检索时的平均精度均值MAP。从表1可以看出,对于两种检索任务,在三种哈希编码长度下,深度跨模态哈希检索方法DCMH、PRDH和本发明方法均能取得比浅层跨模态哈希检索方法SCM和SMFH更好的检索性能。这说明使用深度学习技术学习用于生成二进制哈希编码的深度特征是有益的。从表1中还可以看出,对于Img2Txt和Txt2Img检索任务,在三种哈希编码长度下,本发明方法的跨模态检索性能均优于DCMH和PRDH方法。这说明本发明方法是有效的跨模态哈希检索方法。

[0128] 表1各方法在Pascal VOC 2007数据集上的MAP

[0129]

任务	方法	哈希编码长度		
		16 bits	32 bits	64 bits
Img2Txt	SCM	0.3829	0.3952	0.4094
	SMFH	0.4079	0.4123	0.4204
	DCMH	0.6971	0.7029	0.7052
	PRDH	0.7057	0.7082	0.7112
	本方法	0.7129	0.7190	0.7247
Txt2Img	SCM	0.3741	0.3798	0.3863

[0130]

	SMFH	0.4154	0.4335	0.4378
	DCMH	0.7052	0.6983	0.7105
	PRDH	0.6880	0.6925	0.6973
	本方法	0.7215	0.7293	0.7324

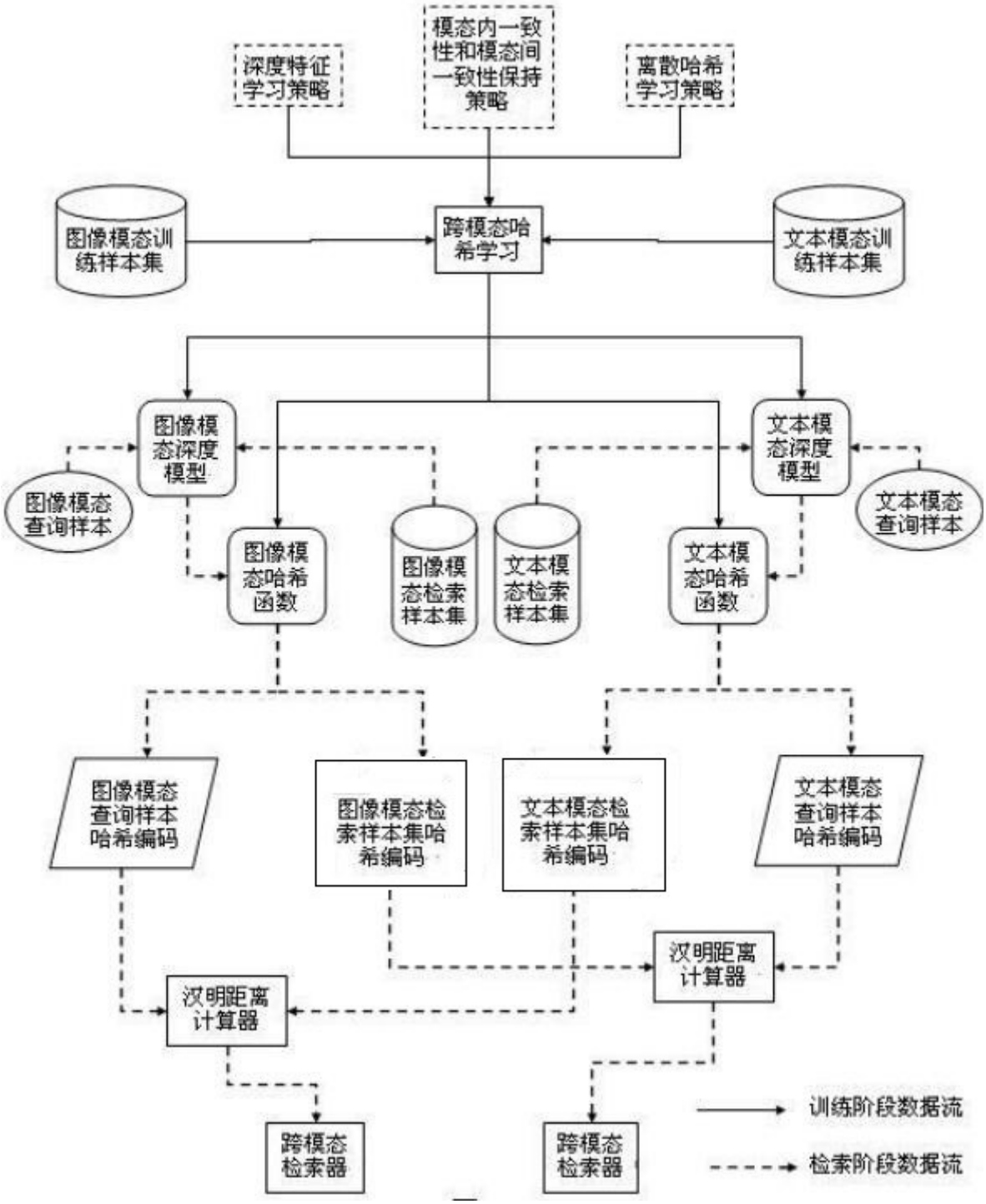

模态内一致性和模态间一致性保持策略
深度特征学习策略
离散哈希学习策略
图像模态训练样本集
跨模态哈希学习
文本模态训练样本集
图像模态深度模型
文本模态深度模型
图像模态查询样本
文本模态查询样本
图像模态哈希函数
图像模态检索样本集
文本模态检索样本集
文本模态哈希函数
图像模态查询样本哈希编码
图像模态检索样本集哈希编码
文本模态检索样本集哈希编码
文本模态查询样本哈希编码
汉明距离计算器
汉明距离计算器
跨模态检索器
跨模态检索器
训练阶段数据流
检索阶段数据流

图1