



(12) 发明专利

(10) 授权公告号 CN 109271486 B

(45) 授权公告日 2021. 11. 26

(21) 申请号 201811097048.3

(22) 申请日 2018.09.19

(65) 同一申请的已公布的文献号  
申请公布号 CN 109271486 A

(43) 申请公布日 2019.01.25

(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 杨茂保 孙丽 董小刚  
尧时茂 王玉伟 邓安远 邓长寿

(74) 专利代理机构 南昌新天下专利商标代理有  
限公司 36115

代理人 谢德珍

(51) Int. Cl.  
G06F 16/33 (2019.01)

(56) 对比文件

CN 106844518 A, 2017.06.13

CN 107346328 A, 2017.11.14

CN 108334574 A, 2018.07.27

CN 107256271 A, 2017.10.17

CN 106777318 A, 2017.05.31

US 2016358043 A1, 2016.12.08

王科. 语义提升和矩阵分解在跨模哈希检索中的应用研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2016, (第10期), I138-439.

李苗 等. 基于解析模态分解法的桥梁动态应变监测数据温度影响的分离.《振动与冲击》.2013, 第31卷(第21期), 6-10, 29.

审查员 杨平勇

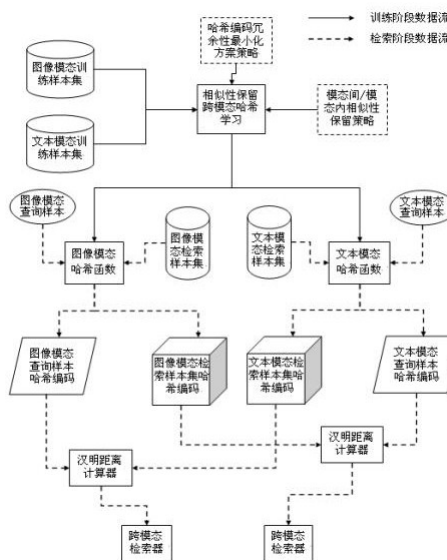
权利要求书3页 说明书13页 附图1页

(54) 发明名称

一种相似性保留跨模态哈希检索方法

(57) 摘要

一种相似性保留跨模态哈希检索方法,该方法包括以下步骤:(1)基于相似性保留策略的目标函数构建;(2)目标函数的求解;(3)生成查询样本和检索样本集中的样本二进制哈希编码;(4)计算查询样本到检索样本集中各个样本的汉明距离;(5)使用跨模态检索器完成对查询样本的检索。本发明方法在进行哈希学习时不仅可以充分保留模态间样本的相似性,还可以充分保留模态内样本的相似性,从而使学习得到的汉明空间具有更强的鉴别能力,更有利于完成跨模态检索。



1. 一种相似性保留跨模态哈希检索方法, 假设 $n$ 个对象 $O = \{o_i\}_{i=1}^n$ 在图像模态和文本模态的特征分别为 $X^{(1)} = [x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}] \in \mathbb{R}^{d_1 \times n}$ 和 $X^{(2)} = [x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}] \in \mathbb{R}^{d_2 \times n}$ , 其中,  $d_1$ 和 $d_2$ 分别表示图像模态和文本模态特征向量的维度,  $x_i^{(1)} \in \mathbb{R}^{d_1}$ 和 $x_i^{(2)} \in \mathbb{R}^{d_2}$ 分别表示第 $i$ 个对象在图像模态和文本模态的特征; 同时假设图像模态和文本模态的特征向量都是经过零中心化预处理的, 即满足 $\sum_{i=1}^n x_i^{(1)} = 0$ ,  $\sum_{i=1}^n x_i^{(2)} = 0$ ; 假设由 $n$ 个对象的类别标记构成的标记矩阵为 $L = [l_1, l_2, \dots, l_n] \in \{0, 1\}^{m \times n}$ , 其中,  $l_i$  ( $i = 1, 2, \dots, n$ ) 表示第 $i$ 个对象的类别标记信息,  $m$ 为类别数量; 假设跨模态相似性矩阵为 $S$ , 其元素 $S_{ij}$ 表示图像模态中第 $i$ 个样本与文本模态中第 $j$ 个样本的相似性; 如果图像模态中第 $i$ 个样本与文本模态中第 $j$ 个样本是相似的, 则 $S_{ij} = 1$ , 否则 $S_{ij} = 0$ ; 其特征在在于, 包括以下步骤:

(1) 基于相似性保留策略的目标函数构建: 使用基于模态间相似性保留策略和模态内相似性保留策略设计的目标函数得到 $n$ 个对象图像模态和文本模态特征数据在汉明空间中的二进制哈希编码 $U$ 和 $V$ , 分别对应于图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ , 以及两个系数矩阵 $W_1$ 和 $W_2$ ;

(2) 目标函数的求解: 鉴于目标函数的非凸性质, 通过交替求解的方式得到目标函数的解 $U, V, P_1, P_2, W_1$ 和 $W_2$ , 即交替地求解如下四个子问题: 固定 $U, V, W_1$ 和 $W_2$ , 求解 $P_1$ 和 $P_2$ ; 固定 $U, V, P_1$ 和 $P_2$ , 求解 $W_1$ 和 $W_2$ ; 固定 $V, P_1, P_2, W_1$ 和 $W_2$ , 求解 $U$ ; 固定 $U, P_1, P_2, W_1$ 和 $W_2$ , 求解 $V$ ;

(3) 生成查询样本和检索样本集中的样本二进制哈希编码: 基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ , 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) 计算查询样本到检索样本集中各个样本的汉明距离: 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) 使用跨模态检索器完成对查询样本的检索: 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。

2. 根据权利要求1所述的一种相似性保留跨模态哈希检索方法, 其特征在在于, 所述的步骤(1)中基于模态间相似性保留策略和模态内相似性保留策略设计的目标函数形式如下:

$$\begin{aligned}
 \min_{U, V, P_1, P_2, W_1, W_2} J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) \\
 & + \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 + \|P_1\|_F^2 + \|P_2\|_F^2 \right) \\
 & + \frac{\beta}{2} \left( \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \left\| \frac{1}{n} V V^T - I \right\|_F^2 \right) + \frac{\eta}{2} \left( \|U \mathbf{1}_{n \times 1}\|_F^2 + \|V \mathbf{1}_{n \times 1}\|_F^2 \right) \\
 \text{s.t. } & U \in \{-1, +1\}^{c \times n}, V \in \{-1, +1\}^{c \times n}
 \end{aligned} \quad (1)$$

其中,  $\alpha, \beta, \gamma$  和 $\eta$ 为非负的平衡因子,  $c$ 为二进制哈希编码的长度,  $I$ 为单位矩阵,  $\mathbf{1}_{n \times 1}$ 表示元素全部为1的列向量,  $\Theta_{ij} = \frac{\lambda}{c} u_i^T v_j$ ,  $\lambda > 0$ 是一个可调节的比例因子,  $u_i$ 为图像模态第 $i$ 个样本的二进制哈希编码,  $v_j$ 为文本模态第 $j$ 个样本的二进制哈希编码,  $\|\cdot\|_F$ 表示矩阵的Frobenius范数,  $(\cdot)^T$ 表示矩阵的转置运算。

3. 根据权利要求2所述的一种相似性保留跨模态哈希检索方法,其特征在于,所述的步骤(2)中通过交替求解的方式得到目标函数的解 $U$ 、 $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,具体为,交替地求解如下四个子问题:

(1) 固定 $U$ 、 $V$ 、 $W_1$ 和 $W_2$ ,求解 $P_1$ 和 $P_2$ ;当固定二进制哈希编码 $U$ 和 $V$ ,以及系数矩阵 $W_1$ 和 $W_2$ 后,公式(1)中的目标函数简化为关于哈希投影矩阵 $P_1$ 和 $P_2$ 的子问题,即:

$$\min_{P_1, P_2} J \equiv \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) + \frac{\gamma}{2} \left( \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (2)$$

(2) 固定 $U$ 、 $V$ 、 $P_1$ 和 $P_2$ ,求解 $W_1$ 和 $W_2$ ;当固定二进制哈希编码 $U$ 和 $V$ 、以及哈希投影矩阵 $P_1$ 和 $P_2$ 后,公式(1)中的目标函数简化为关于系数矩阵 $W_1$ 和 $W_2$ 的子问题,即:

$$\min_{W_1, W_2} J \equiv \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 \right) \quad (3)$$

(3) 固定 $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解 $U$ ;当固定文本模态二进制哈希编码 $V$ ,哈希投影矩阵 $P_1$ 和 $P_2$ ,以及系数矩阵 $W_1$ 和 $W_2$ 后,公式(1)中的目标函数简化为关于图像模态二进制哈希编码 $U$ 的子问题,即:

$$\begin{aligned} \min_U J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|U - P_1^T X^{(1)}\|_F^2 \\ & + \frac{\alpha}{2} \|L - W_1^T U\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \frac{\eta}{2} \|U \mathbf{1}_{n \times 1}\|_F^2 \end{aligned} \quad (4)$$

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

(4) 固定 $U$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解 $V$ ;当固定图像模态二进制哈希编码 $U$ ,哈希投影矩阵 $P_1$ 和 $P_2$ ,以及系数矩阵 $W_1$ 和 $W_2$ 后,公式(1)中的目标函数简化为关于文本模态二进制哈希编码 $V$ 的子问题,即:

$$\begin{aligned} \min_V J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|V - P_2^T X^{(2)}\|_F^2 \\ & + \frac{\alpha}{2} \|L - W_2^T V\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} V V^T - I \right\|_F^2 + \frac{\eta}{2} \|V \mathbf{1}_{n \times 1}\|_F^2 \end{aligned} \quad (5)。$$

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

4. 根据权利要求1所述的一种相似性保留跨模态哈希检索方法,其特征在于,所述的步骤(3)中基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,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 $\tilde{x}^{(1)} \in \mathfrak{R}^{d_1}$ ,文本模态的一个查询样本的特征向量为 $\tilde{x}^{(2)} \in \mathfrak{R}^{d_2}$ ,图像模态检索样本集中样本的特征为 $\tilde{X}^{(1)} = [\tilde{x}_1^{(1)}, \tilde{x}_2^{(1)}, \dots, \tilde{x}_n^{(1)}] \in \mathfrak{R}^{d_1 \times \tilde{n}}$ ,文本模态检索样本集中样本的特征为 $\tilde{X}^{(2)} = [\tilde{x}_1^{(2)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_n^{(2)}] \in \mathfrak{R}^{d_2 \times \tilde{n}}$ ,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的查询样本的二进制哈希编码,以及图像模态和文本模态的检索样本集中样本的二进制哈希编码分别为: $\tilde{u} = P_1^T \tilde{x}^{(1)}$ ,  $\tilde{v} = P_2^T \tilde{x}^{(2)}$ ,  $\tilde{U} = P_1^T \tilde{X}^{(1)}$ 和 $\tilde{V} = P_2^T \tilde{X}^{(2)}$ ,其中,

$$\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_{\tilde{n}}], \quad \tilde{u}_i = P_1^T \tilde{x}_i^{(1)}, \quad \tilde{V} = [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{n}}], \quad \tilde{v}_i = P_2^T \tilde{x}_i^{(2)}, \quad i = 1, 2, \dots, \tilde{n}。$$

5. 根据权利要求4所述的一种相似性保留跨模态哈希检索方法,其特征在於,所述的步骤(4)中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  $D_{\tilde{u}\tilde{v}} = \tilde{u}^T \tilde{V}$  计算图像模态的查询样本到文本模态检索样本集中每个样本的汉明距离,使用公式  $D_{\tilde{v}\tilde{u}} = \tilde{v}^T \tilde{U}$  计算文本模态的查询样本到图像模态检索样本集中每个样本的汉明距离。

6. 根据权利要求5所述的一种相似性保留跨模态哈希检索方法,其特征在於,所述的步骤(5)中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,当用图像检索文本时,对汉明距离  $D_{\tilde{u}\tilde{v}}$  按照从小到大的顺序进行排序,当用文本检索图像时,对汉明距离  $D_{\tilde{v}\tilde{u}}$  按照从小到大的顺序进行排序,然后,在检索样本集中取前K个最小距离对应的样本作为检索结果。

## 一种相似性保留跨模态哈希检索方法

### 技术领域

[0001] 本发明涉及一种相似性保留跨模态哈希检索方法。

### 背景技术

[0002] 在当今社会的各行各业中,大量用户积累了海量的用户数据(比如搜索引擎 Chrome 拥有的数据量超过 100PB),并且数据量还在以指数级的趋势增长,大数据时代接踵而至。大数据在互联网金融、医疗、教育、军事和交通等行业中都具有非常重要的作用,比如,将大数据与机器学习技术结合,能够为金融投资、市场决策等提供可靠的依据。如今的大数据具有以下特性:(1) 体积大,数据量以 PB 为单位;(2) 维度高,数据特征具有数以千计的维度;(3) 模态多,数据种类多、形态多样化,包括图像、文本、音频和视频等形式。大数据的这些特点给机器学习带来了严峻的挑战。面对这种现状,如何合理地利用这些大数据,从中提取出有价值的信息,并为实际的项目工作提供依据是亟待解决的问题。

[0003] 信息检索技术可以为用户检索出有价值的信息,在信息检索领域,相似性搜索是一个研究热点,而近似最近邻搜索 (Approximate Nearest Neighbor, ANN) 因其搜索速度快而备受关注。ANN 搜索方法主要包括树方法和哈希学习方法,两种方法各具特点。具体地,基于树的方法具有如下特性:(1) 递归地划分数据,分而治之;(2) 查询时间复杂度为  $O(\log n)$ ;(3) 随着数据维度的增加,基于树的 ANN 搜索性能会逐渐的下降;(4) 需要存储树结构,存储开销很大;(5) 运行的时候,需要保存原始数据,增加内存的开销。哈希学习方法则具有比较优秀的特性,包括:(1) 数据库中每一项都用二进制串表示,极大地降低数据存储量和内存空间;(2) 查询时间复杂度为常量  $O(1)$  或亚线性。因此,哈希学习方法在实际中得到广泛地应用。

[0004] 跨模态哈希主要用于解决多模态数据之间的相互检索问题,例如,用图像搜索文字,或者用文字搜索图像等问题。跨模态哈希算法需要对不同模态的数据进行哈希编码,生成紧凑的二进制串,然后再进行不同模态数据之间的相互检索。跨模态哈希算法不仅需要考虑同一模态数据之间的关联性,而且还需要考虑不同模态数据之间的关联性。近些年不少跨模态检索哈希方法相继被提出。例如, Bronstein 等人提出了跨模态相似性搜索哈希 (Cross Modality Similarity Search Hashing, CMSSH) 方法, CMSSH 方法将每一位二进制编码对应的哈希函数看成一个弱分类器,通过 AdaBoost 提升算法进行哈希函数的学习。Kuma 等人提出了跨视图哈希 (Cross-View Hashing, CVH) 方法, CVH 方法通过最小化语义相似性与汉明距离之差,为不同的模态数据学习各自的哈希函数。Song 等人提出了媒体间哈希 (Inter-Media Hashing, IMH) 方法, IMH 方法通过保持媒体内与媒体间的一致性去寻找公共的汉明空间,然后再使用线性回归算法为不同的模态学习各自对应的哈希函数。Ding 等人提出了协同矩阵分解哈希 (Collective Matrix Factorization Hashing, CMFH) 方法, CMFH 方法通过协同矩阵分解为不同的模态学习共同的语义表示,然后通过量化方法生成统一的二进制哈希编码。Zhu 等人提出了线性跨模态哈希 (Linear Cross-Modal Hashing, LCMH) 方法, LCMH 方法分别对各个模态内的数据使用 K-means 聚类算法,产生 K 个聚类点,根

据模态内的数据点与这K个聚类点之间的距离重构数据的特征空间,再经过特征值分解求取特征向量,得到各个模态所对应的哈希函数。Zhou等人提出了潜在语义稀疏哈希(Latent Semantics Sparse Hashing, LSSH)方法, LSSH方法结合稀疏编码与矩阵分解技术,为不同模态的特征学习共同的潜在语义表示,然后通过迭代优化算法对目标函数进行最优化求解。Zhang等人基于语义关联最大化(Semantic Correlation Maximization, SCM)方法,通过最大化语义关联性,完成哈希函数的学习,提出了特征分解法SCM\_orth和序列学习法SCM\_seq。Lin等人提出了语义保留哈希(Semantics Preserving Hashing, SePH)方法, SePH方法通过最小化K-L散度,将相似性矩阵转换成概率分布的计算,对各个样本对应的二进制哈希编码串进行概率估算,然后通过核函数回归算法学习各个模态所对应的哈希函数。

[0005] 在将图像模态和文本模态的数据从原始特征空间映射到其它特征空间时,原始数据的一些特征会不可避免地受到损失。对于基于哈希学习的跨模态检索,当将图像模态和文本模态的数据从原始特征空间映射到汉明空间时,原始数据中鉴别信息的有效保留和挖掘,对于有效完成跨模态检索任务具有至关重要的作用。对于图像模态和文本模态的样本来说,不同模态的样本的相似性关系,以及同一模态的样本的相似性关系,是影响跨模态检索的关键因素。不少已有的跨模态哈希学习方法,对于模态间和模态内样本的相似性关系还没有很好地进行处理,有的方法只注重保留模态间样本的相似性关系,有的方法只注重保留模态内样本的相似性关系。这会使得学习得到的汉明空间的鉴别性能受到不利影响。此外,不少方法对哈希编码每一位上所蕴含信息的冗余问题没有进行充分地考虑,使得学习得到的哈希编码不仅具有冗余性还缺少足够的鉴别能力。因此,在跨模态哈希学习中同时注重保留模态间和模态内样本的相似性关系,并使哈希编码每一位上的信息冗余尽可能地小,对于有效地促进跨模态检索性能的提升具有十分重要的意义。

## 发明内容

[0006] 本发明其目的就在于提供一种相似性保留跨模态哈希检索方法,解决了现有的不少方法对模态内和模态间样本的相似性保留不充分,并且对哈希编码每位上的冗余信息消减不充分的问题,使得学习得到的哈希编码具有良好的鉴别能力。

[0007] 为实现上述目的而采取的技术方案是,一种相似性保留跨模态哈希检索方法,假设n个对象 $O = \{o_i\}_{i=1}^n$ 在图像模态和文本模态的特征分别为 $X^{(1)} = [x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}] \in \mathbb{R}^{d_1 \times n}$ 和 $X^{(2)} = [x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}] \in \mathbb{R}^{d_2 \times n}$ ,其中, $d_1$ 和 $d_2$ 分别表示图像模态和文本模态特征向量的维度, $x_i^{(1)} \in \mathbb{R}^{d_1}$ 和 $x_i^{(2)} \in \mathbb{R}^{d_2}$ 分别表示第i个对象在图像模态和文本模态的特征;同时假设图像模态和文本模态的特征向量都是经过零中心化预处理的,即满足 $\sum_{i=1}^n x_i^{(1)} = 0$ ,  $\sum_{i=1}^n x_i^{(2)} = 0$ ;假设由n个对象的类别标记构成的标记矩阵为 $L = [l_1, l_2, \dots, l_n] \in \{0, 1\}^{m \times n}$ ,其中, $l_i$  ( $i=1, 2, \dots, n$ )表示第i个对象的类别标记信息, $m$ 为类别数量;假设跨模态相似性矩阵为 $S$ ,其元素 $S_{i,j}$ 表示图像模态中第i个样本与文本模态中第j个样本的相似性;如果图像模态中第i个样本与文本模态中第j个样本是相似的(至少属于一个相同的类别),则 $S_{i,j}=1$ ,否则 $S_{i,j}=0$ ;本发明方法包括以下步骤:

[0008] (1) 基于相似性保留策略的目标函数构建:使用基于模态间相似性保留策略和模

态内相似性保留策略设计的目标函数得到n个对象图像模态和文本模态特征数据在汉明空间中的二进制哈希编码U和V,分别对应于图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,以及两个系数矩阵 $W_1$ 和 $W_2$ ;

[0009] (2) 目标函数的求解:鉴于目标函数的非凸性质,通过交替更新的方式得到目标函数的解U、V、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,即交替地求解如下四个子问题:固定U、V、 $W_1$ 和 $W_2$ ,求解 $P_1$ 和 $P_2$ ;固定U、V、 $P_1$ 和 $P_2$ ,求解 $W_1$ 和 $W_2$ ;固定V、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解U;固定U、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解V;

[0010] (3) 生成查询样本和检索样本集中的样本二进制哈希编码:基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,为查询样本和检索样本集中的样本生成二进制哈希编码;

[0011] (4) 计算查询样本到检索样本集中各个样本的汉明距离: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;

[0012] (5) 使用跨模态检索器完成对查询样本的检索: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。

[0013] 所述的步骤(1)中基于模态间相似性保留策略和模态内相似性保留策略设计的目标函数形式如下:

$$\begin{aligned}
 \min_{U,V,P_1,P_2,W_1,W_2} J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) \\
 & + \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 + \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (1) \\
 & + \frac{\beta}{2} \left( \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \left\| \frac{1}{n} V V^T - I \right\|_F^2 \right) + \frac{\eta}{2} \left( \|U \mathbf{1}_{n \times 1}\|_F^2 + \|V \mathbf{1}_{n \times 1}\|_F^2 \right) \\
 \text{s.t. } & U \in \{-1, +1\}^{c \times n}, V \in \{-1, +1\}^{c \times n}
 \end{aligned}$$

[0015] 其中, $\alpha$ 、 $\beta$ 、 $\gamma$ 和 $\eta$ 为非负的平衡因子, $c$ 为二进制哈希编码的长度, $I$ 为单位矩阵,

$\mathbf{1}_{n \times 1}$ 表示元素全部为1的列向量,  $\Theta_{ij} = \frac{\lambda}{c} u_i^T v_j$ ,  $\lambda > 0$ 是一个可调节的比例因子, $u_i$ 为图像模态第*i*个样本的二进制哈希编码, $v_j$ 为文本模态第*j*个样本的二进制哈希编码, $\|\cdot\|_F$ 表示矩阵的Frobenius范数, $(\cdot)^T$ 表示矩阵的转置运算。

[0016] 所述的步骤(2)中通过交替求解的方式得到目标函数的解U、V、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,具体为,交替地求解如下四个子问题:

[0017] (1) 固定U、V、 $W_1$ 和 $W_2$ ,求解 $P_1$ 和 $P_2$ ;当固定二进制哈希编码U和V,以及系数矩阵 $W_1$ 和 $W_2$ 后,公式(1)中的目标函数简化为关于哈希投影矩阵 $P_1$ 和 $P_2$ 的子问题,即:

$$\min_{P_1,P_2} J \equiv \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) + \frac{\gamma}{2} \left( \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (2)$$

[0019] (2) 固定U、V、 $P_1$ 和 $P_2$ ,求解 $W_1$ 和 $W_2$ ;当固定二进制哈希编码U和V、以及哈希投影矩阵 $P_1$ 和 $P_2$ 后,公式(1)中的目标函数简化为关于系数矩阵 $W_1$ 和 $W_2$ 的子问题,即:

$$\min_{W_1,W_2} J \equiv \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 \right) \quad (3)$$

[0021] (3) 固定V、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解U;当固定文本模态二进制哈希编码V,哈希投影矩阵 $P_1$ 和 $P_2$ ,以及系数矩阵 $W_1$ 和 $W_2$ 后,公式(1)中的目标函数简化为关于图像模态二进制哈希编

码U的子问题,即:

$$\begin{aligned}
 \min_U J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|U - P_1^T X^{(1)}\|_F^2 \\
 [0022] \quad & + \frac{\alpha}{2} \|L - W_1^T U\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \frac{\eta}{2} \|U \mathbf{1}_{n \times 1}\|_F^2 \quad (4) \\
 \text{s.t. } & U \in \{-1, +1\}^{c \times n}
 \end{aligned}$$

[0023] (4) 固定U、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ , 求解V; 当固定图像模态二进制哈希编码U, 哈希投影矩阵 $P_1$ 和 $P_2$ , 以及系数矩阵 $W_1$ 和 $W_2$ 后, 公式(1)中的目标函数简化为关于文本模态二进制哈希编码V的子问题, 即:

$$\begin{aligned}
 \min_V J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|V - P_2^T X^{(2)}\|_F^2 \\
 [0024] \quad & + \frac{\alpha}{2} \|L - W_2^T V\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} V V^T - I \right\|_F^2 + \frac{\eta}{2} \|V \mathbf{1}_{n \times 1}\|_F^2 \quad (5) \\
 \text{s.t. } & V \in \{-1, +1\}^{c \times n}
 \end{aligned}$$

[0025] 所述的步骤(3)中基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ , 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 $\tilde{x}^{(1)} \in \mathcal{R}^{d_1}$ , 文本模态的一个查询样本的特征向量为 $\tilde{x}^{(2)} \in \mathcal{R}^{d_2}$ , 图像模态检索样本集中样本的特征为 $\tilde{X}^{(1)} = [\tilde{x}_1^{(1)}, \tilde{x}_2^{(1)}, \dots, \tilde{x}_{\tilde{n}}^{(1)}] \in \mathcal{R}^{d_1 \times \tilde{n}}$ , 文本模态检索样本集中样本的特征为 $\tilde{X}^{(2)} = [\tilde{x}_1^{(2)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_{\tilde{n}}^{(2)}] \in \mathcal{R}^{d_2 \times \tilde{n}}$ , 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的查询样本的二进制哈希编码, 以及图像模态和文本模态的检索样本集中样本的二进制哈希编码分别为:  $\tilde{u} = P_1^T \tilde{x}^{(1)}$ ,  $\tilde{v} = P_2^T \tilde{x}^{(2)}$ ,  $\tilde{U} = P_1^T \tilde{X}^{(1)}$ 和 $\tilde{V} = P_2^T \tilde{X}^{(2)}$ , 其中,  $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_{\tilde{n}}]$ ,  $\tilde{u}_i = P_1^T \tilde{x}_i^{(1)}$ ,  $\tilde{V} = [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{n}}]$ ,  $\tilde{v}_i = P_2^T \tilde{x}_i^{(2)}$ ,  $i = 1, 2, \dots, \tilde{n}$ 。

[0026] 所述的步骤(4)中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 $D_{\tilde{u}\tilde{v}} = \tilde{u}^T \tilde{v}$ 计算图像模态的查询样本到文本模态检索样本集中每个样本的汉明距离, 使用公式 $D_{\tilde{v}\tilde{u}} = \tilde{v}^T \tilde{u}$ 计算文本模态的查询样本到图像模态检索样本集中每个样本的汉明距离。

[0027] 所述的步骤(5)中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, 当用图像检索文本时, 对汉明距离 $D_{\tilde{u}\tilde{v}}$ 按照从小到大的顺序进行排序, 当用文本检索图像时, 对汉明距离 $D_{\tilde{v}\tilde{u}}$ 按照从小到大的顺序进行排序, 然后, 在检索样本集中取前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. 本发明方法在进行哈希学习时不仅可以充分保留模态间样本的相似性,还可以充分保留模态内样本的相似性,从而使学习得到的汉明空间具有更强的鉴别能力,更有利于完成跨模态检索;

[0031] 2. 本发明方法针对哈希编码的冗余性进行了充分考虑并通过实施正交约束来最小化哈希编码每一位上的冗余,使得学习得到的哈希编码能够包含更多的鉴别信息,进而使得跨模态检索性能得到有效提升。

## 附图说明

[0032] 以下结合附图对本发明作进一步详述。

[0033] 图1为本发明提出的相似性保留跨模态哈希检索方法工作流程图。

## 具体实施方式

[0034] 一种相似性保留跨模态哈希检索方法,假设 $n$ 个对象 $O = \{o_i\}_{i=1}^n$ 在图像模态和文本模态的特征分别为 $X^{(1)} = [x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}] \in \mathbb{R}^{d_1 \times n}$ 和 $X^{(2)} = [x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}] \in \mathbb{R}^{d_2 \times n}$ ,其中, $d_1$ 和 $d_2$ 分别表示图像模态和文本模态特征向量的维度, $x_i^{(1)} \in \mathbb{R}^{d_1}$ 和 $x_i^{(2)} \in \mathbb{R}^{d_2}$ 分别表示第 $i$ 个对象在图像模态和文本模态的特征;同时假设图像模态和文本模态的特征向量都是经过零中心化预处理的,即满足 $\sum_{i=1}^n x_i^{(1)} = 0$ ,  $\sum_{i=1}^n x_i^{(2)} = 0$ ;假设由 $n$ 个对象的类别标记构成的标记矩阵为 $L = [l_1, l_2, \dots, l_n] \in \{0, 1\}^{m \times n}$ ,其中, $l_i$  ( $i=1, 2, \dots, n$ )表示第 $i$ 个对象的类别标记信息, $m$ 为类别数量;假设跨模态相似性矩阵为 $S$ ,其元素 $S_{ij}$ 表示图像模态中第 $i$ 个样本与文本模态中第 $j$ 个样本的相似性;如果图像模态中第 $i$ 个样本与文本模态中第 $j$ 个样本是相似的(至少属于一个相同的类别),则 $S_{ij}=1$ ,否则 $S_{ij}=0$ ;如图1所示,包括以下步骤:

[0035] (1) 基于相似性保留策略的目标函数构建:使用基于模态间相似性保留策略和模态内相似性保留策略设计的目标函数得到 $n$ 个对象图像模态和文本模态特征数据在汉明空间中的二进制哈希编码 $U$ 和 $V$ ,分别对应于图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,以及两个系数矩阵 $W_1$ 和 $W_2$ ;

[0036] (2) 目标函数的求解:鉴于目标函数的非凸性质,通过交替求解的方式得到目标函数的解 $U$ 、 $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,即交替地求解如下四个子问题:固定 $U$ 、 $V$ 、 $W_1$ 和 $W_2$ ,求解 $P_1$ 和 $P_2$ ;固定 $U$ 、 $V$ 、 $P_1$ 和 $P_2$ ,求解 $W_1$ 和 $W_2$ ;固定 $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解 $U$ ;固定 $U$ 、 $P_1$ 、 $P_2$ 、 $W_1$ 和 $W_2$ ,求解 $V$ ;

[0037] (3) 生成查询样本和检索样本集中的样本二进制哈希编码:基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,为查询样本和检索样本集中的样本生成二进制哈希编码;

[0038] (4) 计算查询样本到检索样本集中各个样本的汉明距离: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;

[0039] (5) 使用跨模态检索器完成对查询样本的检索: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。

[0040] 所述的步骤(1)中基于模态间相似性保留策略和模态内相似性保留策略设计的目标函数形式如下:

$$\begin{aligned}
\min_{U, V, P_1, P_2, W_1, W_2} J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) \\
& + \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 + \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (1) \\
& + \frac{\beta}{2} \left( \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \left\| \frac{1}{n} V V^T - I \right\|_F^2 \right) + \frac{\eta}{2} \left( \|U \mathbf{1}_{n \times 1}\|_F^2 + \|V \mathbf{1}_{n \times 1}\|_F^2 \right)
\end{aligned}$$

[0041]

$$\text{s.t. } U \in \{-1, +1\}^{c \times n}, V \in \{-1, +1\}^{c \times n}$$

[0042] 其中,  $\alpha$ 、 $\beta$ 、 $\gamma$  和  $\eta$  为非负的平衡因子,  $c$  为二进制哈希编码的长度,  $I$  为单位矩阵,

$\mathbf{1}_{n \times 1}$  表示元素全部为1的列向量,  $\Theta_{ij} = \frac{\lambda}{c} u_i^T v_j$ ,  $\lambda > 0$  是一个可调节的比例因子,  $u_i$  为图像模态

第  $i$  个样本的二进制哈希编码,  $v_j$  为文本模态第  $j$  个样本的二进制哈希编码,  $\|\cdot\|_F$  表示矩阵的Frobenius范数,  $(\cdot)^T$  表示矩阵的转置运算。

[0043] 所述的步骤(2)中通过交替求解的方式得到目标函数的解  $U$ 、 $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 具体为, 交替地求解如下四个子问题:[0044] (1) 固定  $U$ 、 $V$ 、 $W_1$  和  $W_2$ , 求解  $P_1$  和  $P_2$ ; 当固定二进制哈希编码  $U$  和  $V$ , 以及系数矩阵  $W_1$  和  $W_2$  后, 公式(1)中的目标函数简化为关于哈希投影矩阵  $P_1$  和  $P_2$  的子问题, 即:

$$\min_{P_1, P_2} J \equiv \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) + \frac{\gamma}{2} \left( \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (2)$$

[0046] (2) 固定  $U$ 、 $V$ 、 $P_1$  和  $P_2$ , 求解  $W_1$  和  $W_2$ ; 当固定二进制哈希编码  $U$  和  $V$ 、以及哈希投影矩阵  $P_1$  和  $P_2$  后, 公式(1)中的目标函数简化为关于系数矩阵  $W_1$  和  $W_2$  的子问题, 即:

$$\min_{W_1, W_2} J \equiv \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 \right) \quad (3)$$

[0048] (3) 固定  $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 求解  $U$ ; 当固定文本模态二进制哈希编码  $V$ , 哈希投影矩阵  $P_1$  和  $P_2$ , 以及系数矩阵  $W_1$  和  $W_2$  后, 公式(1)中的目标函数简化为关于图像模态二进制哈希编码  $U$  的子问题, 即:

$$\begin{aligned}
\min_U J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|U - P_1^T X^{(1)}\|_F^2 \\
& + \frac{\alpha}{2} \|L - W_1^T U\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \frac{\eta}{2} \|U \mathbf{1}_{n \times 1}\|_F^2 \quad (4)
\end{aligned}$$

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

[0050] (4) 固定  $U$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 求解  $V$ ; 当固定图像模态二进制哈希编码  $U$ , 哈希投影矩阵  $P_1$  和  $P_2$ , 以及系数矩阵  $W_1$  和  $W_2$  后, 公式(1)中的目标函数简化为关于文本模态二进制哈希编码  $V$  的子问题, 即:

$$\begin{aligned}
\min_V J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|V - P_2^T X^{(2)}\|_F^2 \\
& + \frac{\alpha}{2} \|L - W_2^T V\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} V V^T - I \right\|_F^2 + \frac{\eta}{2} \|V \mathbf{1}_{n \times 1}\|_F^2 \quad (5)
\end{aligned}$$

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

[0052] 所述的步骤(3)中基于求解得到的图像模态和文本模态的哈希投影矩阵 $P_1$ 和 $P_2$ ,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 $\tilde{x}^{(1)} \in \mathfrak{R}^{d_1}$ ,文本模态的一个查询样本的特征向量为 $\tilde{x}^{(2)} \in \mathfrak{R}^{d_2}$ ,图像模态检索样本集中样本的特征为 $\tilde{X}^{(1)} = [\tilde{x}_1^{(1)}, \tilde{x}_2^{(1)}, \dots, \tilde{x}_{\tilde{n}}^{(1)}] \in \mathfrak{R}^{d_1 \times \tilde{n}}$ ,文本模态检索样本集中样本的特征为 $\tilde{X}^{(2)} = [\tilde{x}_1^{(2)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_{\tilde{n}}^{(2)}] \in \mathfrak{R}^{d_2 \times \tilde{n}}$ ,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的查询样本的二进制哈希编码,以及图像模态和文本模态的检索样本集中样本的二进制哈希编码分别为: $\tilde{u} = P_1^T \tilde{x}^{(1)}$ ,  $\tilde{v} = P_2^T \tilde{x}^{(2)}$ ,  $\tilde{U} = P_1^T \tilde{X}^{(1)}$ 和 $\tilde{V} = P_2^T \tilde{X}^{(2)}$ ,其中, $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_{\tilde{n}}]$ ,  $\tilde{u}_i = P_1^T \tilde{x}_i^{(1)}$ ,  $\tilde{V} = [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{n}}]$ ,  $\tilde{v}_i = P_2^T \tilde{x}_i^{(2)}$ ,  $i = 1, 2, \dots, \tilde{n}$ 。

[0053] 所述的步骤(4)中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 $D_{\tilde{u}\tilde{v}} = \tilde{u}^T \tilde{v}$ 计算图像模态的查询样本到文本模态检索样本集中每个样本的汉明距离,使用公式 $D_{\tilde{v}\tilde{u}} = \tilde{v}^T \tilde{u}$ 计算文本模态的查询样本到图像模态检索样本集中每个样本的汉明距离。

[0054] 所述的步骤(5)中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,当用图像检索文本时,对汉明距离 $D_{\tilde{u}\tilde{v}}$ 按照从小到大的顺序进行排序,当用文本检索图像时,对汉明距离 $D_{\tilde{v}\tilde{u}}$ 按照从小到大的顺序进行排序,然后,在检索样本集中取前K个最小距离对应的样本作为检索结果。

[0055] 具体实施过程主要包括以下步骤:

[0056] (1) 基于相似性保留策略的目标函数构建

[0057] 在本发明方法中,跨模态哈希学习的目的是利用图像模态和文本模态的特征数据 $X^{(1)}$ 和 $X^{(2)}$ ,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 $f^{(1)}(x^{(1)}) \in \{-1, +1\}^{c \times 1}$ 和 $f^{(2)}(x^{(2)}) \in \{-1, +1\}^{c \times 1}$ ,其中, $c$ 是二进制哈希编码可调节的长度。假设 $U = [u_1, u_2, \dots, u_n] \in \{-1, +1\}^{c \times n}$ 和 $V = [v_1, v_2, \dots, v_n] \in \{-1, +1\}^{c \times n}$ 是利用 $n$ 个对象的图像模态和文本模态的特征数据,以及相应的哈希函数生成的哈希空间中的二进制哈希编码,其中, $u_i$ 和 $v_i$ 分别表示第 $i$  ( $i = 1, 2, \dots, n$ )个对象在图像模态和文本模态的哈希编码。为了使哈希空间中的二进制哈希编码 $U$ 和 $V$ 具有良好的鉴别能力,对于哈希编码 $U$ 和 $V$ 来说,希望其能够保留 $S$ 中的相似信息,即若 $S_{ij} = 1$ ,则 $u_i$ 与 $v_j$ 、 $u_j$ 与 $v_i$ 的汉明距离尽量小,反之两者的距离尽量大。

[0058] 为便于说明,仅公式化描述 $u_i$ 与 $v_j$ 的关系, $u_j$ 与 $v_i$ 的关系可以类似地进行公式化描述。对于成对的二进制哈希编码 $\{u_i, v_j\}$ ,基于两者的内积定义哈希编码 $u_i$ 与 $v_j$ 之间的相似性关系,如公式(1)所示:

$$[0059] \quad \Theta_{ij} = \frac{\lambda}{c} \langle u_i, v_j \rangle = \frac{\lambda}{c} u_i^T v_j, \quad (1)$$

[0060] 其中, $\lambda > 0$ 是一个可调节的比例因子, $c$ 是预先设定的哈希编码长度, $\langle \cdot, \cdot \rangle$ 表示向量的内积。使用Sigmoid函数将 $\Theta_{ij}$ 从原始区间投影到 $(0, 1)$ 范围,可得:

$$[0061] \quad A_{ij} = \sigma(\Theta_{ij}) = \frac{1}{1 + e^{-\Theta_{ij}}} \quad (2)$$

[0062] 基于 $A_{ij}$ 定义跨模态相似矩阵S的后验概率,可得:

$$[0063] \quad p(S|U, V) = \prod_{i=1}^n \prod_{j=1}^n p(S_{ij}|U, V) = \prod_{i=1}^n \prod_{j=1}^n (S_{ij} A_{ij} + (1 - S_{ij})(1 - A_{ij})) \quad (3)$$

[0064] 根据概率论中似然函数估计法,将公式(3)的负对数表示为 $J_{\bar{1}}$ :

$$J_{\bar{1}} = -\log p(S|U, V) = \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \text{const}, \quad (4)$$

[0065] 其中, const表示常量。

[0066] 最小化公式(4)可以实现在图像模态和文本模态的哈希编码U和V中保留跨模态相似性。具体来说,从公式(4)可以看出,若 $S_{ij}=1$ ,则 $\Theta_{ij}$ 需要尽可能地大,即, $u_i$ 与 $v_j$ 的内积需要尽量大,也就是说,二进制哈希码 $u_i$ 与 $v_j$ 需要尽量相同,它们之间的汉明距离需要尽量小;反之,若 $S_{ij}=0$ ,则 $\Theta_{ij}$ 需要尽量小,二进制哈希码 $u_i$ 与 $v_j$ 之间的汉明距离需要尽可能地小。

[0067] 一个有效的跨模态检索方法不仅需要考虑模态间相似性,同时还需要考虑模态内的近邻结构,因此,模态内相似性也需要进行保留。对于单个模态来说,该模态的哈希编码是该模态原始特征向量的某种变换。哈希编码模态内相似性保留问题可以作为分类问题来处理,即,最优的哈希编码也能够很好地用于完成分类。假设将图像模态特征 $X^{(1)}$ 和文本模态特征 $X^{(2)}$ 映射为哈希编码U和V的投影矩阵分别为 $P_1 \in \mathbb{R}^{d_1 \times c}$ 和 $P_2 \in \mathbb{R}^{d_2 \times c}$ ,对哈希编码U和V分类的系数矩阵分别为 $W_1 \in \mathbb{R}^{c \times l}$ 和 $W_2 \in \mathbb{R}^{c \times l}$ 。基于 $l_2$ 损失,最小化如下的函数可以达到保留模态内相似性的目的:

$$[0068] \quad J_{\bar{2}} = \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) + \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) \\ + \left( \|W_1\|_F^2 + \|W_2\|_F^2 + \|P_1\|_F^2 + \|P_2\|_F^2 \right) \quad (5) \\ \text{s.t. } U \in \{-1, +1\}^{c \times n}, V \in \{-1, +1\}^{c \times n}$$

[0069] 应用于跨模态检索任务的哈希编码在满足能够保留模态间和模态内相似性的同时,还希望它具有以下特性:

[0070] (1) 独立性。若将哈希编码的每一位看作为一个属性,则希望属性之间的冗余尽可能地小,也就是说,希望位与位之间是相互独立的。该特性的公式化描述如公式(6)所示:

$$[0071] \quad UU^T = nI, VV^T = nI, \quad (6)$$

[0072] 其中,I为单位矩阵。

[0073] (2) 平衡性。也就是说,希望每一位哈希编码是+1和-1的概率相等,各为50%。这一约束可以最大化每一位提供的信息。该特性的公式化描述如公式(7)所示:

$$[0074] \quad U \mathbf{1}_{n \times 1} = 0, V \mathbf{1}_{n \times 1} = 0, \quad (7)$$

[0075] 其中, $\mathbf{1}_{n \times 1}$ 表示元素全部为1的列向量。

[0076] 综合以上分析,相似性保留跨模态哈希检索方法的整体目标函数设计为:

$$\begin{aligned}
[0077] \quad \min_{U, V, P_1, P_2, W_1, W_2} J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) \\
& + \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 + \|P_1\|_F^2 + \|P_2\|_F^2 \right), \quad (8) \\
& + \frac{\beta}{2} \left( \left\| \frac{1}{n} U U^T - I \right\|_F^2 + \left\| \frac{1}{n} V V^T - I \right\|_F^2 \right) + \frac{\eta}{2} \left( \|U \mathbf{1}_{n \times 1}\|_F^2 + \|V \mathbf{1}_{n \times 1}\|_F^2 \right) \\
\text{s.t. } & U \in \{-1, +1\}^{c \times n}, V \in \{-1, +1\}^{c \times n}
\end{aligned}$$

[0078] 其中,  $\alpha$ 、 $\beta$ 、 $\gamma$  和  $\eta$  为非负的平衡因子。

[0079] (2) 目标函数的求解

[0080] 目标函数公式 (8) 中包含六个待求解的变量, 即: 图像模态和文本模态的哈希编码  $U$  和  $V$ , 图像模态和文本模态的哈希投影矩阵  $P_1$  和  $P_2$ , 系数矩阵  $W_1$  和  $W_2$ 。公式 (8) 中的目标函数对于六个待求解的变量是非凸的, 因此, 无法同时得到六个待求解的变量的解析解。公式 (8) 中待求解的未知变量可以通过交替地求解如下四个子问题得到解, 即: 固定  $U$ 、 $V$ 、 $W_1$  和  $W_2$ , 求解  $P_1$  和  $P_2$ ; 固定  $U$ 、 $V$ 、 $P_1$  和  $P_2$ , 求解  $W_1$  和  $W_2$ ; 固定  $V$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 求解  $U$ ; 固定  $U$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 求解  $V$ 。

[0081] (a) 固定  $U$ 、 $V$ 、 $W_1$  和  $W_2$ , 求解  $P_1$  和  $P_2$

[0082] 当固定二进制哈希编码  $U$  和  $V$ , 以及系数矩阵  $W_1$  和  $W_2$  后, 公式 (8) 中的目标函数简化为关于哈希投影矩阵  $P_1$  和  $P_2$  的子问题, 即:

$$[0083] \quad \min_{P_1, P_2} J \equiv \frac{1}{2} \left( \|U - P_1^T X^{(1)}\|_F^2 + \|V - P_2^T X^{(2)}\|_F^2 \right) + \frac{\gamma}{2} \left( \|P_1\|_F^2 + \|P_2\|_F^2 \right). \quad (9)$$

[0084] 公式 (9) 中的问题为标准的偏最小二乘回归问题。分别对  $J$  关于  $P_1$  和  $P_2$  求偏导数并令导数等于 0, 可得:

$$[0085] \quad \frac{\partial J}{\partial P_1} = \left( X^{(1)} X^{(1)T} - X^{(1)} U^T \right) P_1 + \gamma P_1 = 0, \quad (10)$$

$$[0086] \quad \frac{\partial J}{\partial P_2} = \left( X^{(2)} X^{(2)T} - X^{(2)} V^T \right) P_2 + \gamma P_2 = 0. \quad (11)$$

[0087] 对公式 (10) 和公式 (11) 进行简单推导, 可得:

$$[0088] \quad P_1 = \left( X^{(1)} X^{(1)T} + \gamma I \right)^{-1} X^{(1)} U^T, \quad (12)$$

$$[0089] \quad P_2 = \left( X^{(2)} X^{(2)T} + \gamma I \right)^{-1} X^{(2)} V^T, \quad (13)$$

[0090] 其中,  $(\cdot)^{-1}$  表示矩阵的逆矩阵。

[0091] (b) 固定  $U$ 、 $V$ 、 $P_1$  和  $P_2$ , 求解  $W_1$  和  $W_2$

[0092] 当固定二进制哈希编码  $U$  和  $V$ 、以及哈希投影矩阵  $P_1$  和  $P_2$  后, 公式 (8) 中的目标函数简化为关于系数矩阵  $W_1$  和  $W_2$  的子问题, 即:

$$[0093] \quad \min_{W_1, W_2} J \equiv \frac{\alpha}{2} \left( \|L - W_1^T U\|_F^2 + \|L - W_2^T V\|_F^2 \right) + \frac{\gamma}{2} \left( \|W_1\|_F^2 + \|W_2\|_F^2 \right). \quad (14)$$

[0094] 公式 (14) 中的问题也是标准的偏最小二乘回归问题。分别对  $J$  关于  $W_1$  和  $W_2$  求偏导数并令导数等于 0, 可得:

[0095] 
$$\frac{\partial J}{\partial W_1} = \alpha(UU^T - UL^T)W_1 + \gamma W_1 = 0, \quad (15)$$

[0096] 
$$\frac{\partial J}{\partial W_2} = \alpha(VV^T - VL^T)W_2 + \gamma W_2 = 0. \quad (16)$$

[0097] 对公式(15)和公式(16)进行简单推导,可得:

[0098] 
$$W_1 = (UU^T + (\gamma I) / \alpha)^{-1}UL^T, \quad (17)$$

[0099] 
$$W_2 = (VV^T + (\gamma I) / \alpha)^{-1}VL^T. \quad (18)$$

[0100] (c) 固定V、P<sub>1</sub>、P<sub>2</sub>、W<sub>1</sub>和W<sub>2</sub>,求解U

[0101] 当固定文本模态二进制哈希编码V,哈希投影矩阵P<sub>1</sub>和P<sub>2</sub>,以及系数矩阵W<sub>1</sub>和W<sub>2</sub>后,公式(8)中的目标函数简化为关于图像模态二进制哈希编码U的子问题,即:

[0102] 
$$\begin{aligned} \min_U J(U) = & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|U - P_1^T X^{(1)}\|_F^2 \\ & + \frac{\alpha}{2} \|L - W_1^T U\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} UU^T - I \right\|_F^2 + \frac{\eta}{2} \|U \mathbf{1}_{n \times 1}\|_F^2. \quad (19) \\ \text{s.t. } & U \in \{-1, +1\}^{c \times n} \end{aligned}$$

[0103] 为了方便求解,本发明方法将离散哈希变量U松弛为连续变量进行求解,因此,公式(19)中的目标函数转化为:

[0104] 
$$\begin{aligned} \min_U J(U) = & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|U - P_1^T X^{(1)}\|_F^2 \\ & + \frac{\alpha}{2} \|L - W_1^T U\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} UU^T - I \right\|_F^2 + \frac{\eta}{2} \|U \mathbf{1}_{n \times 1}\|_F^2. \quad (20) \end{aligned}$$

[0105] 对J(U)关于U求导数可得:

[0106] 
$$\begin{aligned} \frac{\partial J(U)}{\partial U} = & \frac{\lambda}{c} \sum_{i=1}^n \sum_{j=1}^n \left( \sigma(\Theta_{ij}) - S_{ij} \right) V + (U - P_1^T X^{(1)}) \\ & + \alpha W_1 (W_1^T U - L) + \beta \frac{1}{n} \left( \frac{1}{n} UU^T - I \right) U + \eta U \mathbf{1}_{n \times 1} \mathbf{1}_{n \times 1}^T. \quad (21) \end{aligned}$$

[0107] 为了求得U的最优解,使用梯度下降法迭代更新U,即,通过U(t+1) = U(t) + ΔU(t)迭代更新U.具体地,根据泰勒展开式,有:

[0108] 
$$J(U + \Delta U) \approx J(U) + \Delta U^T \frac{\partial J(U)}{\partial U}. \quad (22)$$

[0109] 于是,欲满足J(U + ΔU) < J(U),可以选择:

[0110] 
$$\Delta U = -\omega_1 \frac{\partial J(U)}{\partial U}, \quad (23)$$

[0111] 其中,步长ω<sub>1</sub>是一个预先定义的比较小的常数.当连续变量U求得之后,使用公式

$U = \text{sign}(U)$  得到离散的哈希变量  $U$ , 其中,  $\text{sign}(\cdot)$  为符号函数, 即: 当  $x \geq 0$  时,  $\text{sign}(x) = +1$ , 当  $x < 0$  时,  $\text{sign}(x) = -1$ 。

[0112] (d) 固定  $U$ 、 $P_1$ 、 $P_2$ 、 $W_1$  和  $W_2$ , 求解  $V$

[0113] 当固定图像模态二进制哈希编码  $U$ , 哈希投影矩阵  $P_1$  和  $P_2$ , 以及系数矩阵  $W_1$  和  $W_2$  后, 公式 (8) 中的目标函数简化为关于文本模态二进制哈希编码  $V$  的子问题, 即:

$$\begin{aligned} \min_V J \equiv & \sum_{i=1}^n \sum_{j=1}^n \left( \log(1 + e^{\Theta_{ij}}) - S_{ij} \Theta_{ij} \right) + \frac{1}{2} \|V - P_2^T X^{(2)}\|_F^2 \\ [0114] \quad & + \frac{\alpha}{2} \|L - W_2^T V\|_F^2 + \frac{\beta}{2} \left\| \frac{1}{n} V V^T - I \right\|_F^2 + \frac{\eta}{2} \|V \mathbf{1}_{n \times 1}\|_F^2 \quad (24) \\ \text{s.t. } & V \in \{-1, +1\}^{c \times n} \end{aligned}$$

[0115] 类似于对离散变量  $U$  的求解, 同样将离散哈希变量  $V$  松弛为连续变量进行求解。当连续变量  $V$  求得之后, 使用公式  $V = \text{sign}(V)$  得到离散的哈希变量  $V$ 。

[0116] (3) 生成查询样本和检索样本集中的样本二进制哈希编码

[0117] 假设图像模态的一个查询样本的特征向量为  $\tilde{x}^{(1)} \in \mathfrak{R}^{d_1}$ , 文本模态的一个查询样本的特征向量为  $\tilde{x}^{(2)} \in \mathfrak{R}^{d_2}$ , 图像模态检索样本集中样本的特征为  $\tilde{X}^{(1)} = [\tilde{x}_1^{(1)}, \tilde{x}_2^{(1)}, \dots, \tilde{x}_{\tilde{n}}^{(1)}] \in \mathfrak{R}^{d_1 \times \tilde{n}}$ , 文本模态检索样本集中样本的特征为  $\tilde{X}^{(2)} = [\tilde{x}_1^{(2)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_{\tilde{n}}^{(2)}] \in \mathfrak{R}^{d_2 \times \tilde{n}}$ , 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的查询样本的二进制哈希编码, 以及图像模态和文本模态的检索样本集中样本的二进制哈希编码分别为:  $\tilde{u} = P_1^T \tilde{x}^{(1)}$ ,  $\tilde{v} = P_2^T \tilde{x}^{(2)}$ ,  $\tilde{U} = P_1^T \tilde{X}^{(1)}$  和  $\tilde{V} = P_2^T \tilde{X}^{(2)}$ , 其中,  $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_{\tilde{n}}]$ ,  $\tilde{u}_i = P_1^T \tilde{x}_i^{(1)}$ ,  $\tilde{V} = [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{n}}]$ ,  $\tilde{v}_i = P_2^T \tilde{x}_i^{(2)}$ ,  $i = 1, 2, \dots, \tilde{n}$ 。

[0118] (4) 计算查询样本到检索样本集中各个样本的汉明距离

[0119] 对于图像模态的查询样本  $\tilde{x}^{(1)}$ , 使用公式  $D_{\tilde{u}\tilde{V}} = [d_{\tilde{u}\tilde{v}_1}, d_{\tilde{u}\tilde{v}_2}, \dots, d_{\tilde{u}\tilde{v}_{\tilde{n}}}] = \tilde{u}^T \tilde{V} = \tilde{u}^T [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{\tilde{n}}]$  计算图像模态的查询样本  $\tilde{x}^{(1)}$  到文本模态检索样本集中每个样本  $\tilde{x}_i^{(2)}$  ( $i = 1, 2, \dots, \tilde{n}$ ) 的汉明距离。对于文本模态的查询样本  $\tilde{x}^{(2)}$ , 使用公式  $D_{\tilde{v}\tilde{U}} = [d_{\tilde{v}\tilde{u}_1}, d_{\tilde{v}\tilde{u}_2}, \dots, d_{\tilde{v}\tilde{u}_{\tilde{n}}}] = \tilde{v}^T \tilde{U} = \tilde{v}^T [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_{\tilde{n}}]$  计算文本模态的查询样本  $\tilde{x}^{(2)}$  到图像模态检索样本集中每个样本  $\tilde{x}_i^{(1)}$  ( $i = 1, 2, \dots, \tilde{n}$ ) 的汉明距离。

[0120] (5) 使用跨模态检索器完成对查询样本的检索

[0121] 当用图像检索文本时, 对汉明距离  $D_{\tilde{u}\tilde{V}}$  按照从小到大的顺序进行排序, 当用文本检索图像时, 对汉明距离  $D_{\tilde{v}\tilde{U}}$  按照从小到大的顺序进行排序, 然后, 在检索样本集中取前  $K$  个最小距离对应的样本作为检索结果。

[0122] 以下结合具体实验对本发明的有益效果进行说明。

[0123] MIRFLICKR-25K 数据集中包含从 Flickr 网站收集的 25000 幅图像, 并且每一幅图像附加有 24 种文本标签中的若干种文本标签, 因此, 可以认为 MIRFLICKR-25K 数据集是一个多标签数据集。在实验中只选取至少包含 20 个文本标签的样本, 构成 20015 对图像-文本样本。对于每对图像-文本, 每幅图像用 512 维的 GIST 特征向量表示, 文本的特征用 1386 维的词袋

向量表示。在实验中随机选择1000对图像-文本样本用于构造查询样本集,随机选择5000对图像-文本样本用于训练跨模态检索模型。

[0124] 在实验中,用平均精度均值 (MeanAverage Precision,MAP) 衡量跨模态检索方法的性能。要计算MAP需要首先计算平均精度 (Average Precision,AP)。假设某个查询样本在进行跨模态检索时返回R个检索到的样本,对应于这个查询样本的平均精度AP定义为:

$$[0125] \quad AP = \frac{\sum_{r=1}^R P(r)\delta(r)}{\sum_{r=1}^R \delta(r)}。 \quad (25)$$

[0126] 在公式 (25) 中,P (r) 表示前r个被检索到的样本的精度,即前r个被检索到的样本中有多少比例的样本是真正与查询样本相关的。对于 $\delta(r)$ ,当第r个被检索到的样本真正与查询样本相关时, $\delta(r)=1$ ,反之, $\delta(r)=0$ 。当获取了所有查询样本的平均精度AP后,计算这些平均精度AP的均值即可得到平均精度均值MAP。

[0127] 在实验中,本发明方法中的参数 $\alpha$ 、 $\beta$ 、 $\gamma$  和 $\eta$ 采用5折交叉验证确定最优的值。对于其它方法中的参数,参照各个方法文献中所采用的参数设置方式进行参数设置。实验中报告的结果均为10次随机实验结果的平均值。

[0128] 与本发明方法进行对比的方法分别为:典型相关分析 (Canonical Correlation Analysis,CCA) 方法、跨视图哈希 (Cross-View Hashing,CVH) 方法、媒体间哈希 (Inter-Media Hashing,IMH) 方法和潜在语义稀疏哈希 (Latent Semantic Sparse Hashing,LSSH) 方法。表1总结了本发明提出的方法和相关方法在MIRFLICKR-25K数据集上进行跨模态检索时的平均精度均值MAP。在表1中,Img2Txt和Txt2Img分别表示用图像检索文本的任务和用文本检索图像的任务。从表1中可以看出,对于图像检索文本和文本检索图像这两种检索任务,本发明方法的检索性能在四种哈希编码长度下都胜过对比方法。具体来说,与其它对比方法相比,对于Img2Txt任务,本发明方法在16bits、32bits、64bits和128bits上的MAP至少提升了0.0152(=0.3121-0.2969)、0.022(=0.3285-0.3065)、0.0253(=0.3371-0.3118)和0.0196(=0.3442-0.3246);对于Txt2Img任务,本发明方法在16bits、32bits、64bits和128bits上的MAP至少提升了0.0242(=0.3925-0.3683)、0.0278(=0.4257-0.3979)、0.0273(=0.4618-0.4345)和0.0351(=0.4969-0.4618)。这说明本发明提出的相似性保留跨模态哈希检索方法是有效的。

[0129] 表1 MIRFLICKR-25K数据集上各方法的MAP



任务	方法	哈希编码长度			
		16 bits	32 bits	64 bits	128 bits
[0130] Img2Txt	CCA	0.2732	0.2913	0.2982	0.3117
	CVH	0.2842	0.2993	0.3054	0.3135
	IMH	0.2858	0.3038	0.3124	0.3189
	LSSH	0.2969	0.3065	0.3118	0.3246
	本方法	0.3121	0.3285	0.3371	0.3442
Txt2Img	CCA	0.3369	0.3635	0.3989	0.4319
	CVH	0.3472	0.3761	0.4168	0.4417
	IMH	0.3491	0.3802	0.4099	0.4514
	LSSH	0.3683	0.3979	0.4345	0.4618
	本方法	0.3925	0.4257	0.4618	0.4969

[0131] 本发明还包括：模态间样本相似性保留策略，模态内样本相似性保留策略，以及哈希编码冗余性最小化方案。

[0132] 所述的模态间样本相似性保留策略：对于跨模态检索任务来说，在具体的检索过程中需要面对来自不同模态性质差异很大的异质数据，有效地消除不同模态数据的异质性，充分地从不模态数据的复杂关系中挖掘出它们之间本质的联系，能够促进跨模态检索性能的提升。为了从不模态的数据中充分地挖掘鉴别信息，本发明方法基于内积定义不同模态样本的哈希编码的相似性关系，并利用Sigmoid函数将这种相似性关系建模为概率模型，然后基于跨模态相似矩阵的后验概率完成模态间相似性的保留，进而达到从跨模态异质数据中有效挖掘鉴别信息的目的。

[0133] 所述的模态内样本相似性保留策略：对于模态内部的样本来说，样本的标记信息能够有效地反映它们之间的近邻结构和相似性关系。对于单个模态来说，该模态的哈希编码是该模态的原始特征数据从原始特征空间到汉明空间的某种变换。为了使哈希编码在模态内部保留样本之间的相似性，本发明方法借助于样本的标记信息，通过利用完成分类任务的线性回归模型来完成模态内相似性的保留。

[0134] 所述的哈希编码冗余性最小化方案：对于哈希编码来说，若将哈希编码的每一位看作为一个属性，则希望不同属性之间的冗余尽可能地小，也就是说，希望位与位之间能够是相互独立的。本发明方法通过对哈希编码的不同位实施正交约束来达到这个目的。

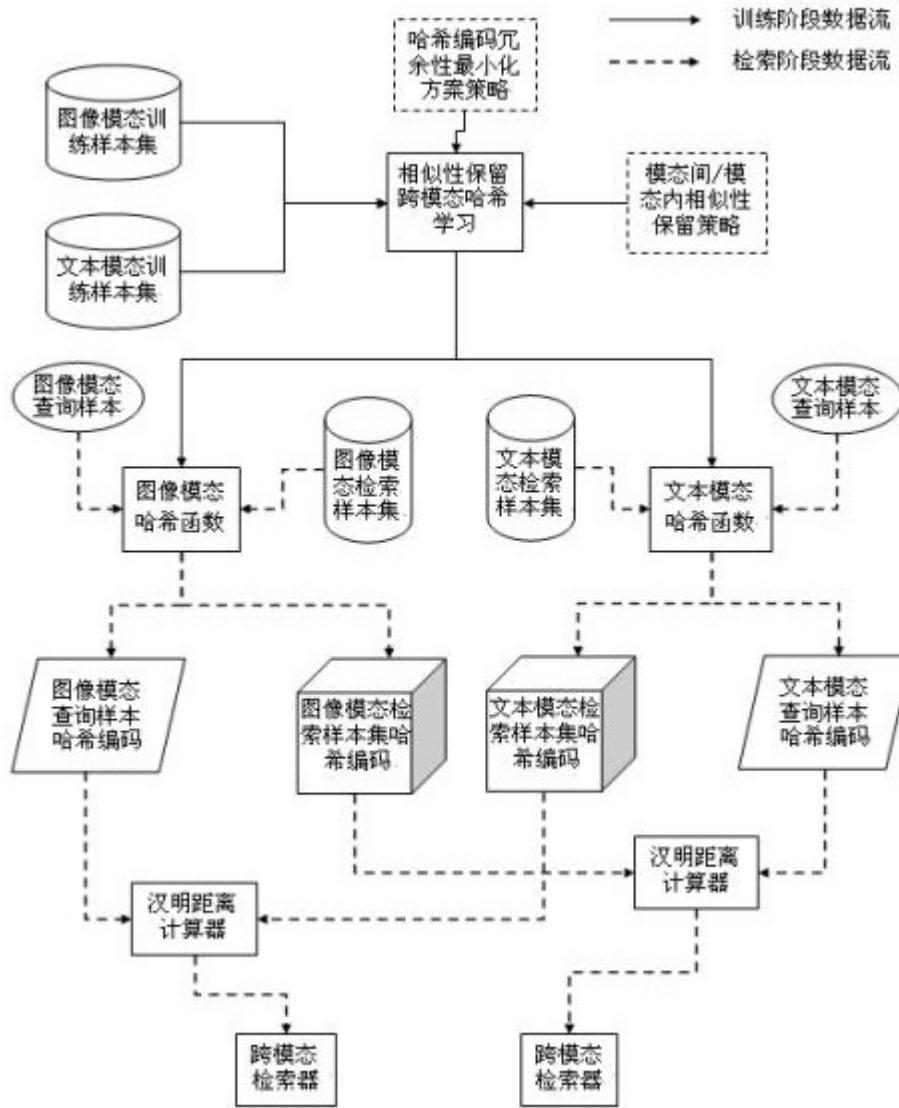


图1