



(12)发明专利

(10)授权公告号 CN 109376261 B

(45)授权公告日 2019.09.24

(21)申请号 201811268589.8

G06K 9/62(2006.01)

(22)申请日 2018.10.29

审查员 王国海

(65)同一申请的已公布的文献号

申请公布号 CN 109376261 A

(43)申请公布日 2019.02.22

(73)专利权人 山东师范大学

地址 250014 山东省济南市历下区文化东路88号

(72)发明人 张化祥 郑顺心 李静 吴泓辰

王琳 孙建德

(74)专利代理机构 济南圣达知识产权代理有限公司

公司 37221

代理人 黄海丽

(51)Int.Cl.

G06F 16/58(2019.01)

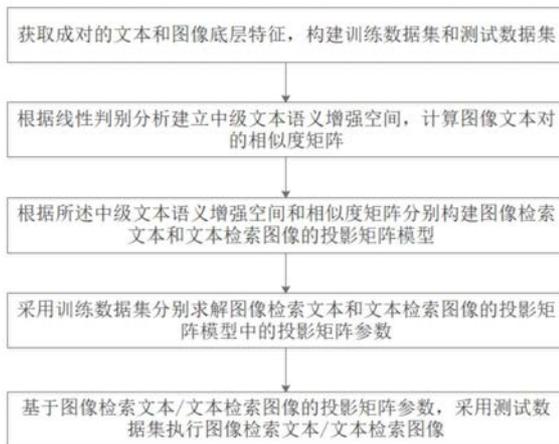
权利要求书2页 说明书9页 附图2页

(54)发明名称

基于中级文本语义增强空间的模态独立检索方法和系统

(57)摘要

本发明公开了一种基于中级文本语义增强空间的模态独立检索方法和系统,所述方法包括:获取成对的文本和图像底层特征,构建训练数据集和测试数据集;根据线性判别分析建立中级文本语义增强空间;计算图像文本对的相似度矩阵;根据所述中级文本语义增强空间和相似度矩阵分别构建图像检索文本和文本检索图像的投影矩阵模型;采用训练数据集分别求解图像检索文本和文本检索图像的投影矩阵模型中的投影矩阵参数;基于图像检索文本/文本检索图像的投影矩阵参数,采用测试数据集执行图像检索文本/文本检索图像。所述方法为图像检索文本和文本检索图像的检索任务分别学习映射矩阵,检索更有针对性,大大提高了检索的准确率。



1. 一种基于中级文本语义增强空间的模态独立检索方法,其特征在于,包括以下步骤:
 获取成对的文本和图像底层特征,构建训练数据集和测试数据集;
 根据线性判别分析建立中级文本语义增强空间;计算图像文本对的相似度矩阵;
 根据所述中级文本语义增强空间和相似度矩阵分别构建图像检索文本和文本检索图像的投影矩阵模型;

采用训练数据集分别求解图像检索文本和文本检索图像的投影矩阵模型中的投影矩阵参数;

基于图像检索文本/文本检索图像的投影矩阵参数,采用测试数据集执行图像检索文本/文本检索图像;

所述计算图像文本对的相似度矩阵包括:

根据语义标签计算图像文本对的相似度矩阵 $S = \{s_{ij}\}_{2n \times 2n}$;

定义 S_{TT} 和 S_{II} 为单媒体相似矩阵,均为零矩阵, S_{TI} 和 S_{IT} 为跨媒体相似矩阵;

定义对角矩阵 $D = \text{Diag}(d_{1,1}, d_{2,2}, \dots, d_{2n,2n})$, $d_{i,i} = \sum_{k=1}^{2n} s_{ik}$;

将 $S = \{s_{ij}\}_{2n \times 2n}$ 表示为 $S = \begin{pmatrix} S_{TT} & S_{TI} \\ S_{IT} & S_{II} \end{pmatrix}$,其中, $S_{IT} = S_{TI}^T$, $S_{II} = S_{TT}$, $S = S^T$,并进行归一

化;最终相似度矩阵 $L = D - S = \begin{pmatrix} L_{TT} & L_{TI} \\ L_{IT} & L_{II} \end{pmatrix}$, $L_{IT} = L_{TI}^T$ 和 $L_{II} = L_{TT}$;

所述图像检索文本投影矩阵模型为:

$$\Omega(V_1, W_1) = \text{tr}((TW_1^T)^T L_{TT} TW_1^T + 2(TW_1^T)^T L_{TI} IV_1^T + (IV_1^T) Y^T L_{II} IV_1^T) + \alpha \|IV_1^T - Y\|_F^2 + \beta \|V_1\|_F^2 + \beta \|W_1\|_F^2 + \text{tr}(W_1 U_w W_1^T) - \lambda \text{tr}(W_1 U_t W_1^T)$$

其中, $V_1 \in \mathbb{R}^{k \times c}$ 为图像的投影矩阵, $W_1 \in \mathbb{R}^{k \times c}$ 为文本的投影矩阵;

所述文本检索图像投影矩阵模型为:

$$\Omega(V_2, W_2) = \text{tr}((TW_2^T)^T L_{TT} TW_2^T + 2(TW_2^T)^T L_{TI} IV_2^T + (IV_2^T) Y^T L_{II} IV_2^T) + \alpha \|TW_2^T - Y\|_F^2 + \beta \|V_2\|_F^2 + \beta \|W_2\|_F^2 + \text{tr}(W_2 U_w W_2^T) - \lambda \text{tr}(W_2 U_t W_2^T)$$

其中 $V_2 \in \mathbb{R}^{k \times c}$ 为图像的投影矩阵, $W_2 \in \mathbb{R}^{k \times c}$ 为文本的投影矩阵;

T 为文本的底层特征矩阵, I 为图像的底层特征矩阵, Y 是标记矩阵, α , β 和 λ 为平衡参数, U_w 为类内散度矩阵, U_t 为总体散度矩阵, $\text{tr}(\cdot)$ 表示矩阵的迹。

2. 如权利要求1所述的一种基于中级文本语义增强空间的模态独立检索方法,其特征在于,所述中级文本语义空间的建立方法包括:

根据每个类文本特征和所有文本特征的均值,计算文本特征的类内散度矩阵和类间散度矩阵;

将所述类内散度矩阵和类间散度矩阵代入线性判别分析公式,得到线性判别分析的目标函数。

3. 如权利要求1所述的一种基于中级文本语义增强空间的模态独立检索方法,其特征在于,定义 s_{ij} 是文本和图像的相似性矩阵,定义 y_i 为第 i 个文本图像对对应的标签矩阵向

量, y_j 为第 j 个文本图像对对应的标签矩阵向量, s_{ij} 可以表示为:

$$s_{ij} = \begin{cases} 1, & \text{when } y_i = y_j \\ -1, & \text{when } y_i \neq y_j \end{cases}。$$

4. 如权利要求1所述的一种基于中级文本语义增强空间的模态独立检索方法, 其特征在于, 所述投影矩阵模型包括:

用于保持图像文本对的关联关系的媒体相关项, 采用相似度矩阵;

用于提升文本特征判别性的正则化项, 采用线性判别分析的目标函数;

用于针对不同的检索任务学习投影矩阵的线性投影项;

用于控制投影矩阵的复杂度的正则化项。

5. 如权利要求1所述的一种基于中级文本语义增强空间的模态独立检索方法, 其特征在于, 所述求解图像检索文本投影矩阵模型或文本检索图像投影矩阵模型包括: 固定文本投影矩阵求解图像投影矩阵; 固定图像投影矩阵求解文本投影矩阵。

6. 一种计算机系统, 包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序, 其特征在于, 所述处理器执行所述程序时实现如权利要求1-5任一项所述的基于中级文本语义增强空间的模态独立检索方法。

7. 一种计算机可读存储介质, 其上存储有计算机程序, 其特征在于, 该程序被处理器执行时实现如权利要求1-5任一项所述的基于中级文本语义增强空间的模态独立检索方法。

基于中级文本语义增强空间的模态独立检索方法和系统

技术领域

[0001] 本公开属于跨媒体检索技术领域,尤其涉及一种基于中级文本语义增强空间的模态独立检索方法和系统。

背景技术

[0002] 随着社会的发展和网络的普及,网络上的多媒体数据,图片、视频、音频等呈现出指数级的增长,许多不同类型的多媒体数据会表达同一种语义,用户需要从这些海量的数据中检索到自己需要的信息。当前机器学习和模式识别理论的飞速发展,硬件的更新换代使得计算速度得到极大的提升,如何实现跨媒体数据的检索已成为亟需解决的问题。

[0003] 跨媒体检索是指提交任何类型的媒体类型的数据,获得不同媒体类型的相同语义的结果。目前的检索方式还主要局限在单一媒体检索,基于关键字的搜索源于文本检索,用户提供关键字,数据库通过关键词关联媒体数据,查询系统再根据关键字查找相关数据并返回信息,如当前主流的搜索引擎百度、搜狗、谷歌等采用的技术都是基于关键字搜索。基于内容的图像检索,一般需要提取图像的特征来描述一幅图像,然后查询系统通过特征来查找特征类似的图像数据。而当前的网络上尤其是社交媒体上充满了大量没有关键字的多媒体数据,类似上述的搜索方法对没有关键字或者跨越多媒体的检索无法有效的查询。

[0004] 由于不同类型的多媒体数据在原始特征的维度和属性的不同会导致跨媒体数据之间存在异构性问题,同一类型但语义不同的多媒体数据会存在原始特征与其高层语义不一致的问题,即“跨媒体异构鸿沟”和“语义鸿沟”。为解决上述问题,在处理图像和文本的检索任务中,很多研究采用子空间学习方法,典型相关性分析(CCA和CFA)通过保持多媒体数据相关性的前提下,将多媒体数据投影到同一维度的子空间中,进而计算跨媒体相似性,以解决“跨媒体异构鸿沟”问题。

[0005] 然而,由于图像中往往包含大量杂乱的信息,在图像特征的提取过程中会存在大量的语义鸿沟,所以文本模态的特征往往比图像特征具有更强的判别性,传统的子空间方法并没有有效地利用文本特征的强判别性来改善检索结果。同时,大多数方法对于不同的检索任务仅仅学习一对映射矩阵,这种映射机制在单一检索任务中并不能达到最高的性能。

发明内容

[0006] 为克服上述现有技术的不足,本公开提供了一种基于中级文本语义增强空间的模态独立检索方法和系统,所述方法为图像检索文本和文本检索图像的检索任务分别学习映射矩阵,检索更有针对性,大大提高了检索的准确率。

[0007] 为实现上述目的,本公开的一个或多个实施例采用如下技术方案:

[0008] 一种基于中级文本语义增强空间的模态独立检索方法,包括以下步骤:

[0009] 获取成对的文本和图像底层特征,构建训练数据集和测试数据集;

[0010] 根据线性判别分析建立中级文本语义增强空间;计算图像文本对的相似度矩阵;

[0011] 根据所述中级文本语义增强空间和相似度矩阵分别构建图像检索文本和文本检索图像的投影矩阵模型；

[0012] 采用训练数据集分别求解图像检索文本和文本检索图像的投影矩阵模型中的投影矩阵参数；

[0013] 基于图像检索文本/文本检索图像的投影矩阵参数,采用测试数据集执行图像检索文本/文本检索图像。

[0014] 进一步地,所述中级文本语义空间的建立方法包括:

[0015] 根据每个类文本特征和所有文本特征的均值,计算文本特征的类内散度矩阵和类间散度矩阵;

[0016] 将所述类内散度矩阵和类间散度矩阵代入线性判别分析公式,得到线性判别分析的目标函数。

[0017] 进一步地,所述计算图像文本对的相似度矩阵包括:

[0018] 根据语义标签计算图像文本对的相似度矩阵 $S = \{s_{ij}\}_{2m \times 2m}$;

[0019] 定义 S_{TT} 和 S_{II} 为单媒体相似矩阵,均为零矩阵, S_{TI} 和 S_{IT} 为跨媒体相似矩阵;

[0020] 定义对角矩阵 $D = \text{Diag}(d_{1,1}, d_{2,2}, \dots, d_{2n,2n})$, $d_{i,i} = \sum_{i=1}^n s_{ii}$;

[0021] 将 $S = \{s_{ij}\}_{2m \times 2m}$ 表示为 $S = \begin{pmatrix} S_{TT} & S_{TI} \\ S_{IT} & S_{II} \end{pmatrix}$,其中, $S_{IT} = S_{TI}^T$, $S_{II} = S_{TT}$, $S = S^T$,并进行

归一化;最终相似度矩阵 $L = D - S = \begin{pmatrix} L_{TT} & L_{TI} \\ L_{IT} & L_{II} \end{pmatrix}$, $L_{IT} = L_{TI}^T$ 和 $L_{II} = L_{TT}$ 。

[0022] 进一步地,

[0023] 定义 s_{ij} 是文本和图像的相似性矩阵,定义 y_i 为第 i 个文本图像对对应的标签矩阵向量, y_j 为第 j 个文本图像对对应的标签矩阵向量, s_{ij} 可以表示为:

[0024]
$$s_{ij} = \begin{cases} 1, & \text{when } y_i = y_j \\ -1, & \text{when } y_i \neq y_j \end{cases}。$$

[0025] 进一步地,所述投影矩阵模型包括:

[0026] 用于保持图像文本对的关联关系的媒体相关项,采用相似度矩阵;

[0027] 用于提升文本特征判别性的正则化项,采用线性判别分析的目标函数;

[0028] 用于针对不同的检索任务学习投影矩阵的线性投影项;

[0029] 用于控制投影矩阵的复杂度的正则化项。

[0030] 进一步地,所述图像检索文本投影矩阵模型为:

[0031]
$$\Omega(V_1, W_1) = \text{tr}((TW_1^T)^T L_{TT} TW_1^T + 2(TW_1^T)^T L_{TI} IV_1^T + (IV_1^T) Y^T L_{II} IV_1^T) + \alpha \|IV_1^T - Y\|_F^2 + \beta \|V_1\|_F^2 + \beta \|W_1\|_F^2 + \text{tr}(W_1 U_w W_1^T) - \lambda \text{tr}(W_1 U_t W_1^T)$$

[0032] 其中, $V_1 \in R^{k \times c}$ 为图像的投影矩阵, $W_1 \in R^{k \times c}$ 为文本的投影矩阵。

[0033] 进一步地,所述文本检索图像投影矩阵模型为:

$$\begin{aligned}
 [0034] \quad \Omega(V_2, W_2) = & \operatorname{tr}((TW_2^T)^T L_{TT} TW_2^T + 2(TW_2^T)^T L_{TI} IV_2^T + (IV_2^T) Y^T L_{II} IV_2^T) + \alpha \|TW_2^T - Y\|_F^2 \\
 & + \beta \|V_2\|_F^2 + \beta \|W_2\|_F^2 + \operatorname{tr}(W_2 U_w W_2^T) - \lambda \operatorname{tr}(W_2 U_l W_2^T)
 \end{aligned}$$

[0035] 其中 $V_2 \in \mathbb{R}^{k \times c}$ 为图像的投影矩阵, $W_2 \in \mathbb{R}^{k \times c}$ 为文本的投影矩阵。

[0036] 进一步地, 所述求解图像检索文本投影矩阵模型或文本检索图像投影矩阵模型包括: 固定文本投影矩阵求解图像投影矩阵; 固定图像投影矩阵求解文本投影矩阵。

[0037] 一个或多个实施例提供了一种计算机系统, 包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序, 所述处理器执行所述程序时实现所述的基于中级文本语义增强空间的模态独立检索方法。

[0038] 一个或多个实施例提供了一种计算机可读存储介质, 其上存储有计算机程序, 该程序被处理器执行时实现所述的基于中级文本语义增强空间的模态独立检索方法。

[0039] 本公开的有益效果

[0040] 本公开通过建立一个中级文本语义增强空间, 充分利用文本特征的高判别性, 并将提高的文本特征的通过距离保持投影迁徙到图像特征中, 使得图像特征中的噪音影响更低, 最后, 在投影过程中根据不同的任务学习不同的投影矩阵, 更有检索的针对性, 大大提高了检索的准确率。

附图说明

[0041] 构成本申请的一部分的说明书附图用来提供对本申请的进一步理解, 本申请的示意性实施例及其说明用于解释本申请, 并不构成对本申请的不当限定。

[0042] 图1为本公开实施例一中基于中级文本语义增强空间的模态独立检索方法的流程图;

[0043] 图2为本公开实施例一中基于中级文本语义增强空间的模态独立检索方法的流程示意图。

[0044] 图3为本公开实施例一中训练投影矩阵模型的流程图。

具体实施方式

[0045] 应该指出, 以下详细说明都是示例性的, 旨在对本申请提供进一步的说明。除非另有指明, 本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的含义。

[0046] 需要注意的是, 这里所使用的术语仅是为了描述具体实施方式, 而非意图限制根据本申请的示例性实施方式。如在这里所使用的, 除非上下文另外明确指出, 否则单数形式也意图包括复数形式, 此外, 还应当理解的是, 当在本说明书中使用术语“包含”和/或“包括”时, 其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

[0047] 在不冲突的情况下, 本申请中的实施例及实施例中的特征可以相互组合。

[0048] 正如背景技术所介绍的, 现有技术中存在忽略文本特征的高判别性, 图像特征的噪音过多, 没有针对不同的任务学习不同的矩阵已达到最优的结果的缺点。本公开为了解决上述问题, 提出了一种基于中级文本语义增强空间的模态独立检索方法。建立一个中级语义空间, 文本特征的判别能力得到进一步增强并且这种判别能力通过共享子空间迁移到

对应的图像特征中,从而提高图像特征的代表能力。具体而言,利用线性判别分析把文本信息投影到中级语义增强空间中,用以增强文本特征的判别能力,通过距离保持投影生成映射矩阵,把增强的文本特征的判别能力迁移到图像特征中。最后,在检索过程中针对不同的跨媒体检索任务学习不同的映射矩阵,因此能够提高检索准确度,并通过实验验证了本公开的有效性。

[0049] 实施例一

[0050] 本实施例公开了一种基于中级文本语义增强空间的模态独立检索方法,如图1所示,该方法包括以下步骤:

[0051] 步骤1:获取成对文本和图像的底层特征,构建文本的训练集和测试集、图像的训练集和测试集;

[0052] 所述步骤1具体包括:

[0053] 步骤1:输入三个数据集,获取每个数据集图像和文本的底层特征,表示为 $\Omega = \{(i_i, t_i)\}_{i=1}^n$, (i_i, t_i) 表示第*i*对表示相同语义的文本和图像特征,*n*是样例个数,数据集中共有*k*个类。设 $I = [i_1, i_2, \dots, i_n] \in \mathbb{R}^{n \times p}$ 为图像的底层特征矩阵,*p*是图像的维度。设 $T = [t_1, t_2, \dots, t_n] \in \mathbb{R}^{n \times q}$ 为文本的底层特征矩阵,*q*是文本的维度。 $Y = [y_1, y_2, \dots, y_n] \in \mathbb{R}^{n \times k}$ 是标记矩阵。设图像的投影矩阵为 $V \in \mathbb{R}^{k \times p}$,文本的投影矩阵为 $W \in \mathbb{R}^{k \times q}$ 。并将数据集中70%样本用于训练,30%样本用于测试。

[0054] 步骤2:建立中级文本语义增强空间;

[0055] 中级文本语义空间的具体建立过程包括:

[0056] 步骤2.1:计算第*j*个类中文本特征的均值,计算所有文本特征的均值,计算文本的类内散度矩阵和类间散度矩阵。

[0057] 设 \bar{m}_j 第*j*类文本特征的均值, \bar{m} 全部类的文本特征的均值,则类内散度矩阵可表述为 $U_w = \sum_{j=1}^c \sum_{y_i=j} \frac{1}{n} (t_i - \bar{m}_j)(t_i - \bar{m}_j)^T$, 总体散度矩阵可以表述为

$$U_t = \sum_{i=1}^n \frac{1}{n} (t_i - \bar{m})(t_i - \bar{m})^T。$$

[0058] 步骤2.2:把步骤2.1得到的矩阵代入线性判别分析(LDA)公式,得到LDA的目标函数。通过线性判别分析(LDA)对文本特征进行降维处理,使得相同类之间更加聚合,不同类之间更加远离,以提高文本特征的判别能力。

[0059] 利用步骤2.1得到的 U_w 和 U_t 可得到线性判别分析表达式:

$$[0060] \quad \min_{U^T=I_k} \frac{\text{tr}(WU_w W^T)}{\text{tr}(WU_t W^T)} \quad (1)$$

[0061] 步骤3:计算文本图像对的相似度矩阵*L*;

[0062] 所述步骤3中,计算图像文本对的相似度矩阵。该系数值越大,说明文本和图像对的相似度越高。

[0063] 步骤3.1:定义 s_{ij} 是文本和图像的相似性矩阵,定义 y_i 为第*i*个文本图像对对应的标签矩阵向量, y_j 为第*j*个文本图像对对应的标签矩阵向量, s_{ij} 可以表示为:

$$[0064] \quad s_{ij} = \begin{cases} 1, & \text{when } y_i = y_j \\ -1, & \text{when } y_i \neq y_j \end{cases} \quad (2)$$

[0065] 步骤3.2:定义跨媒体相似矩阵 S_{TI} 和 S_{IT} ,定义单媒体相似矩阵 S_{TT} 和 S_{II} ,定义包含所有媒体类型的整体相似度矩阵 $S = \{s_{ij}\}_{2m \times 2m}$ 可以表示为:

$$[0066] \quad S = \begin{pmatrix} S_{TT} & S_{TI} \\ S_{IT} & S_{II} \end{pmatrix} \quad (3)$$

[0067] 上述式子中, $S_{IT} = S_{TI} = S_{TT}^T, S_{II} = S_{TT}$ and $S = S^T$ 。利用跨媒体相关性,因此设置 S_{TT} 和 S_{II} 为零矩阵。为了平衡正相关和负相关的影响,将 S 归一化,使 S 中的每个行的和等于零,并将 S 变换为对称矩阵。因此 $S_{IT} = S_{TI} = S_{TT}^T, S_{II} = S_{TT}, S = S^T$,上标 T 为矩阵的转置。

[0068] 步骤3.3:定义对角矩阵 $D = \text{Diag}(d_{1,1}, d_{2,2}, \dots, d_{2n,2n})$, $d_{ii} = \sum_{i=1}^n s_{ii}$ 。定义 $L_{IT} = L_{TI} = L_{TT}^T$ 和 $L_{II} = L_{TT}$,则 L 可以表示为

$$[0069] \quad L = D - S = \begin{pmatrix} L_{TT} & L_{TI} \\ L_{IT} & L_{II} \end{pmatrix} \quad (4)$$

[0070] 步骤4:对图像检索文本和文本检索图像分别构建投影矩阵模型。

[0071] 所述步骤4中,投影矩阵模型的具体建立过程包括:

[0072] 步骤4.1:利用步骤3中得到的相似度矩阵作为跨媒体相关项 $f(V, W)$,用来保持图像文本对的关联关系,

[0073] 综合步骤3.1-3.3, $f(V, W)$ 可以表示为:

$$[0074] \quad f(V, W) = \sum_{i=1}^n \sum_{j=1}^n s_{ij} \|t_i W - i_j V\|_2^2 \quad (5)$$

[0075] 利用公式(4), $f(V, W)$ 可以以矩阵形式重写为:

$$[0076] \quad \begin{aligned} &= \text{tr} \left(\begin{pmatrix} TW \\ IV \end{pmatrix}^T L \begin{pmatrix} TW \\ IV \end{pmatrix} \right) \\ &= \text{tr} \left((TW)^T L_{TT} TW + 2(TW)^T L_{TI} IV + (IV)^T L_{II} IV \right) \end{aligned} \quad (6)$$

[0077] 步骤4.2:利用步骤2.2中的得到的LDA的目标函数作为正则化项 $\lambda l(W, U_w, U_t)$,用来提升文本特征的判别性。

[0078] 步骤4.3:加入一项线性投影项 $\alpha g(V, W)$,用来针对不同的检索任务学习不同的投影矩阵。

[0079] 步骤4.4:加入一项正则化项 $\beta r(V, W)$,控制投影矩阵的复杂度,防止学习过程过拟合。

[0080] 得到如下的关于 V 和 W 的最优化问题,

$$[0081] \quad \min_{V, W} f(V, W) + \alpha g(V, W) + \beta r(V, W) + \lambda l(W, U_w, U_t) \quad (7)$$

[0082] 其中 α, β 和 λ 为平衡参数,取值在(0-1)之间。 $\min_{V, W} f(V, W)$ 为距离保持投影项,保持

图像文本对关系。 $\alpha g(V, W)$ 为线性回归项, 通过学习投影矩阵, 在图像检索文本时将图像的原始特征矩阵投影到高级语义空间, 在文本检索图像时将文本的原始特征矩阵投影到高级语义空间。 $\beta r(V, W)$ 为误差项, 控制投影矩阵的复杂度, 防止过拟合。 $\lambda l(W, U_w, U_t)$ 为线性判别分析项, 对文本特征进行降维处理, 增强文本特征的判别性。

[0083] 步骤5: 对图像检索文本投影矩阵模型和文本检索图像投影矩阵模型分别输入训练集进行训练。

[0084] 所述步骤5中, 具体的训练过程包括:

[0085] 步骤5.1: 输入训练样本的特征矩阵, 语义矩阵。设置迭代过程的步长, 设置收敛条件。初始化投影矩阵, 初始化参数;

[0086] 步骤5.2: 计算中级空间文本投影矩阵和距离保持投影项;

[0087] 步骤5.3: 优化目标函数; 执行图像检索文本或文本检索图像;

[0088] 更新权重, 重复目标函数优化步骤, 直至目标函数收敛。

[0089] 具体地, 所述步骤5.3中, 图像检索文本训练过程如下:

[0090] 假设 x 是一个矩阵, $\text{tr}(x)$ 表示矩阵 x 的迹。图像检索文本的目标函数如下, :

[0091]

$$\begin{aligned} \Omega(V_1, W_1) = & \text{tr}((TW_1^T)^T L_{TT} TW_1^T + 2(TW_1^T)^T L_{TI} IV_1^T + (IV_1^T) Y^T L_{II} IV_1^T) + \alpha \|IV_1^T - Y\|_F^2 \\ & + \beta \|V_1\|_F^2 + \beta \|W_1\|_F^2 + \text{tr}(W_1 U_w W_1^T) - \lambda \text{tr}(W_1 U_t W_1^T) \end{aligned} \quad (8)$$

[0092] 其中 $V_1 \in \mathbb{R}^{k \times c}$ 为图像的投影矩阵, $W_1 \in \mathbb{R}^{k \times c}$ 为文本的投影矩阵, $\text{tr}(\cdot)$ 表示矩阵的迹。

[0093] 求解过程包括:

[0094] 步骤(1-1): 固定 W_1 求 V_1 得到:

$$\Delta_{V_1} \Omega(V_1, W_1) = 2I^T L_{II} V^T + 2I^T L_{TI} TW^T + 2\alpha I^T (IV^T - Y) + 2\beta V \quad (9)$$

[0096] 步骤(1-2): 固定 V_1 求 W_1 得到:

[0097]

$$\Delta_{W_1} \Omega(V_1, W_1) = 2T^T L_{TT} TW^T + 2T^T L_{TI} IV^T + 2\alpha W^T + W_1 U_w^T + W_1 U_w - \lambda(W_1 U_t^T + W_1 U_t) \quad (10)$$

[0098] 步骤(1-3): 重复步骤(1-1) - (1-2), 直到收敛。输出 W_1 和 V_1

[0099] (2) 文本检索图像训练过程如下:

[0100] 文本检索图像的优化过程与图像检索文本的过程基本一致, 在下面主要介绍区别:

[0101] 文本检索图像的目标函数为:

[0102]

$$\begin{aligned} \Omega(V_2, W_2) = & \text{tr}((TW_2^T)^T L_{TT} TW_2^T + 2(TW_2^T)^T L_{TI} IV_2^T + (IV_2^T) Y^T L_{II} IV_2^T) + \alpha \|TW_2^T - Y\|_F^2 \\ & + \beta \|V_2\|_F^2 + \beta \|W_2\|_F^2 + \text{tr}(W_2 U_w W_2^T) - \lambda \text{tr}(W_2 U_t W_2^T) \end{aligned} \quad (12)$$

[0103] 其中 $V_2 \in \mathbb{R}^{k \times c}$ 为图像的投影矩阵, $W_2 \in \mathbb{R}^{k \times c}$ 为文本的投影矩阵。与图像检索文本区别在于该往高层语义空间投影文本特征矩阵。

[0104] 求解过程包括:

[0105] 步骤(2-1):固定 W_2 求 V_2 得到:

$$[0106] \quad \Delta_{V_2} \Omega(V_2, W_2) = 2I^T L_{II} IV^T + 2I^T L_{TI} TW^T + 2\beta V \quad (13)$$

[0107] 步骤(2-2):固定 V_1 求 W_1 得到:

$$[0108] \quad \Delta_{W_2} \Omega(V_2, W_2) = 2T^T L_{TI} TW^T + 2T^T L_{II} IV^T + 2\alpha T^T (TW^T - Y) + \dots \quad (14)$$

$$+ 2\beta W^T + W_2 U_w^T + W_2 U_w - \lambda(W_2 U_t^T + W_2 U_t)$$

[0109] 步骤(2-3):重复步骤(2-1)-(2-2),直到收敛。输出 W_2 和 V_2

[0110] 步骤5.4:输出图像和文本的投影矩阵。

[0111] 步骤6:对不同的模型分别输入不同的测试集数据,测试模型的查准率。

[0112] 所述步骤6中,具体的测试步骤包括:

[0113] 步骤6.1:分别向文本和图像的投影矩阵模型输入测试文本特征和图像特征。

[0114] 步骤6.2:通过计算查询对象和所有被查询对象两两之间的欧式距离,确定最匹配的查询结果。

[0115] 步骤6.3:通过计算平均查准率的均值来体现模型的准确率。

[0116] 输入测试集,进行测试。对检索结果进行排序,得到最终的检索结果,本实施例采用平均AP和精确回忆曲线(PR)被用来评价跨媒体检索的性能。AP值的计算公式如下:

$$[0117] \quad AP = \frac{\sum_{j=1}^R P(j)rel(j)}{\sum_{i=1}^R rel(j)} \quad (11)$$

[0118] 其中的R为查询结果的数量。如果在第j项的样本与查询项的语义相同,则 $rel(j) = 1$,否则为0。 $P(j)$ 是检索的前k个结果的准确率。

[0119] 以Wikipedia数据集、pascal数据集和Websearch数据集上九种跨媒体检索方法的检索平均查准率的均值(mAP)比较(表1-2):

[0120] 表1与现有方法对比实验结果

[0121]

Dataset	Query	PLS	CCA	SM	SCM	GMMFA
Wikipedia	I2T	0.359	0.331	0.368	0.374	0.284
dataset two	T2I	0.351	0.316	0.386	0.392	0.248
	Average	0.355	0.324	0.377	0.383	0.266
Pascal	I2T	0.365	0.379	0.449	0.407	0.373
Sentences	T2I	0.376	0.372	0.433	0.393	0.347
dataset	Average	0.370	0.379	0.441	0.400	0.360
INRIA	I2T	0.193	0.260	0.378	0.354	0.280
-Websearch	T2I	0.260	0.279	0.353	0.308	0.303
dataset	Average	0.227	0.269	0.365	0.331	0.292

[0122] 表2与现有方法对比实验结果(续表1)

	Dataset	Query	GMLDA	JFSSL	MDCR	JLSLR	本方法
	Wikipedia	I2T	0.300	0.392	0.410	0.393	0.438
	dataset two	T2I	0.280	0.381	0.377	0.369	0.395
		Average	0.290	0.386	0.394	0.381	0.417
[0123]	Pascal	I2T	0.408	0.406	0.432	0.454	0.485
	Sentences	T2I	0.387	0.401	0.462	0.455	0.490
	dataset	Average	0.397	0.404	0.447	0.454	0.488
	INRIA	I2T	0.475	0.532	0.470	0.525	0.526
	-Websearch	T2I	0.540	0.562	0.459	0.544	0.552
[0124]	dataset	Average	0.508	0.547	0.465	0.534	0.539

[0125] 从表1中可以看出,本公开取得了最好的异构媒体检索结果,本公开解决了现有技术中存在忽略文本特征的高判别性,图像特征的噪音过多,没有针对不同的任务学习不同的矩阵的缺点,提出了一种基于中级文本语义增强空间的模态独立检索方法。建立一个中级语义空间,文本特征的判别能力得到进一步增强并且这种判别能力通过共享子空间迁移到对应的图像特征中,从而提高图像特征的表征能力。具体而言,利用线性判别分析把文本信息投影到中级语义增强空间中,用以增强文本特征的判别能力,通过距离保持投影生成映射矩阵,把增强的文本特征的判别能力迁移到图像特征中。最后,在检索过程中针对不同的跨媒体检索任务学习不同的映射矩阵,因此能够提高检索准确度,并通过实验验证了本公开的有效性。

[0126] 实施例二

[0127] 本实施例的目的是提供一种计算机系统。

[0128] 一种计算机系统,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现以下步骤,包括:

[0129] 获取成对的文本和图像底层特征,构建训练数据集和测试数据集;

[0130] 根据线性判别分析建立中级文本语义增强空间;计算图像文本对的相似度矩阵;

[0131] 根据所述中级文本语义增强空间和相似度矩阵分别构建图像检索文本和文本检索图像的投影矩阵模型;

[0132] 采用训练数据集分别求解图像检索文本和文本检索图像的投影矩阵模型中的投影矩阵参数;

[0133] 基于图像检索文本/文本检索图像的投影矩阵参数,采用测试数据集执行图像检索文本/文本检索图像。

[0134] 实施例三

[0135] 本实施例的目的是提供一种计算机可读存储介质。

[0136] 一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时执行以下步骤:

[0137] 获取成对的文本和图像底层特征,构建训练数据集和测试数据集;

- [0138] 根据线性判别分析建立中级文本语义增强空间;计算图像文本对的相似度矩阵;
- [0139] 根据所述中级文本语义增强空间和相似度矩阵分别构建图像检索文本和文本检索图像的投影矩阵模型;
- [0140] 采用训练数据集分别求解图像检索文本和文本检索图像的投影矩阵模型中的投影矩阵参数;
- [0141] 基于图像检索文本/文本检索图像的投影矩阵参数,采用测试数据集执行图像检索文本/文本检索图像。
- [0142] 以上实施例二和三中涉及的各步骤与方法实施例一相对应,具体实施方式可参见实施例一的相关说明部分。术语“计算机可读存储介质”应该理解为包括一个或多个指令集的单个介质或多个介质;还应当被理解为包括任何介质,所述任何介质能够存储、编码或承载用于由处理器执行的指令集并使处理器执行本公开中的任一方法。
- [0143] 上述一个或多个实施例具有以下技术效果:
- [0144] 本公开通过建立一个中级文本语义增强空间,充分利用文本特征的高判别性,并将提高的文本特征的通过距离保持投影迁徙到图像特征中,使得图像特征中的噪音影响更低,最后,在投影过程中根据不同的任务学习不同的投影矩阵,更有检索的针对性,大大提高了检索的准确率。
- [0145] 本领域技术人员应该明白,上述本申请的各模块或各步骤可以用通用的计算机装置来实现,可选地,它们可以用计算装置可执行的程序代码来实现,从而,可以将它们存储在存储装置中由计算装置来执行,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。本申请不限制于任何特定的硬件和软件的结合。
- [0146] 以上所述仅为本申请的优选实施例而已,并不用于限制本申请,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。
- [0147] 上述虽然结合附图对本申请的具体实施方式进行了描述,但并非对本申请保护范围的限制,所属领域技术人员应该明白,在本申请的技术方案的基础上,本领域技术人员不需要付出创造性劳动即可做出的各种修改或变形仍在本申请的保护范围以内。

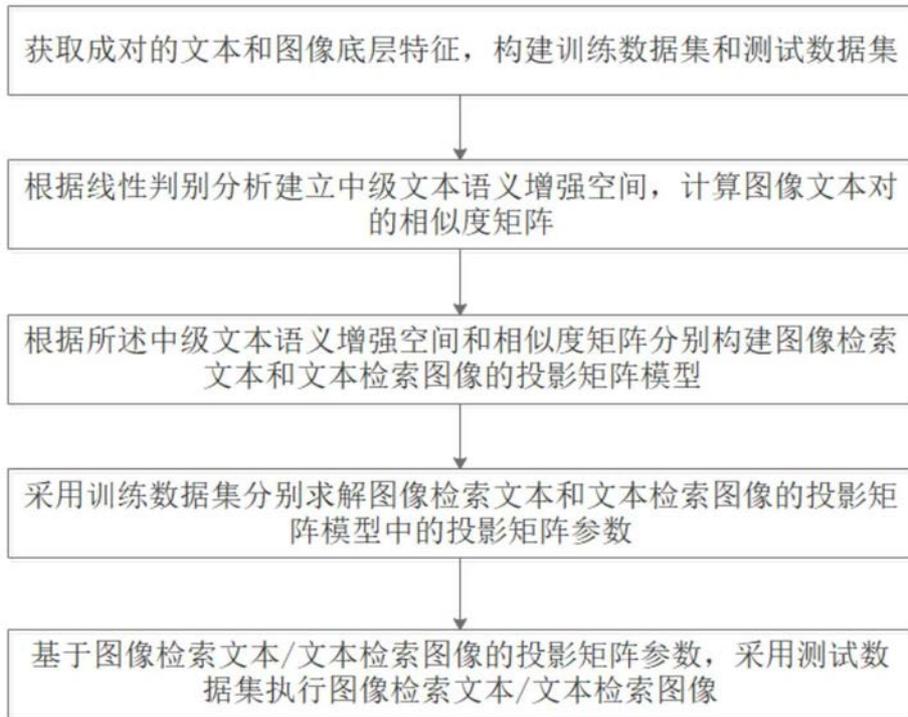


图1

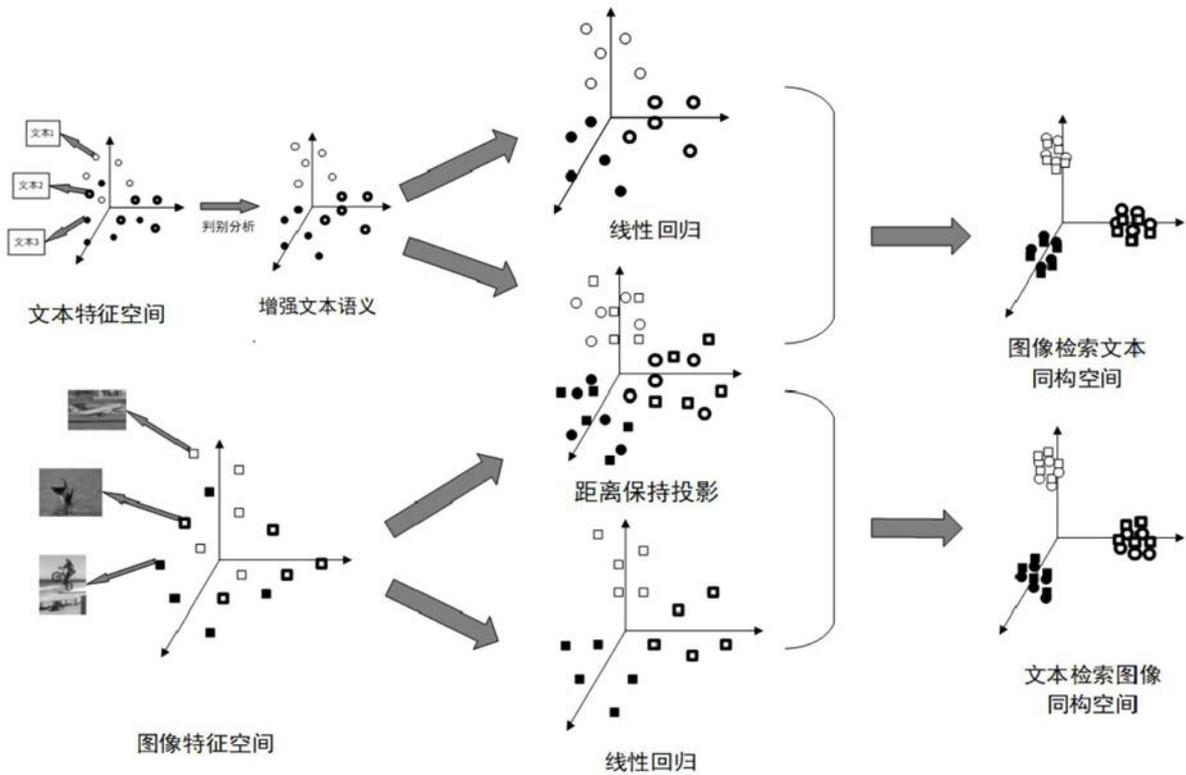


图2

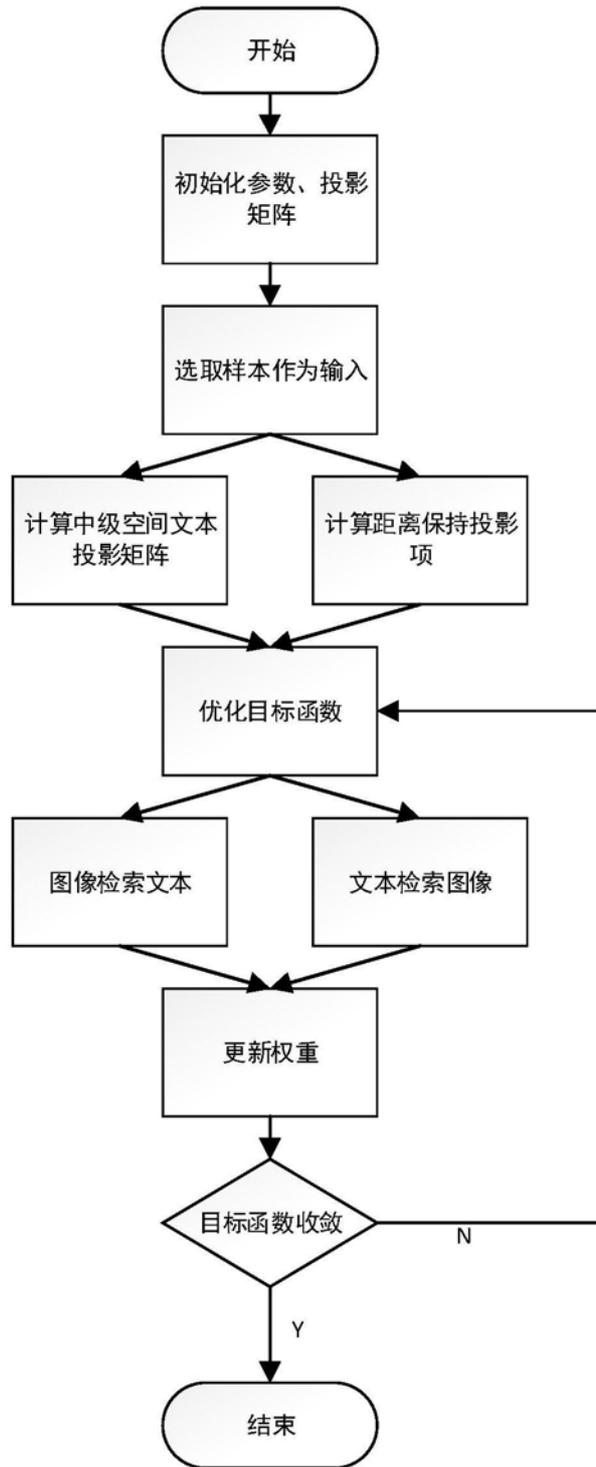


图3