



(12)发明专利申请

(10)申请公布号 CN 105930873 A

(43)申请公布日 2016.09.07

(21)申请号 201610274436.9

G06F 17/30(2006.01)

(22)申请日 2016.04.27

(71)申请人 天津中科智能识别产业技术研究院  
有限公司

地址 300465 天津市滨海新区天津经济技术  
开发区第二大街57号泰达MSD-G1-  
1001单元

(72)发明人 赫然 孙哲南 李志航 梁坚  
曹冬

(74)专利代理机构 天津市三利专利商标代理有  
限公司 12107

代理人 韩新城

(51) Int. Cl.

G06K 9/62(2006.01)

G06K 9/46(2006.01)

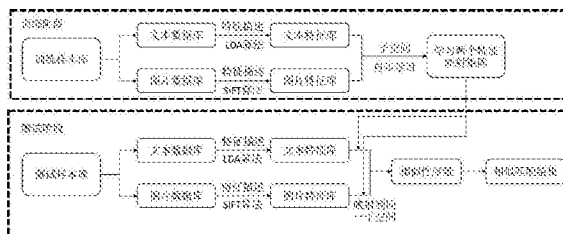
权利要求书3页 说明书5页 附图1页

(54)发明名称

一种基于子空间的自步跨模态匹配方法

(57)摘要

本发明公开一种基于子空间的自步跨模态匹配方法。该方法通过提取数据集中不同模态数据的特征向量;利用子空间自步学习得到训练集中与不同模态对应的不同映射矩阵,利用该映射矩阵将测试集中数据样本的不同模态类别映射到同一空间,使训练集与测试集的数据映射到同一空间,然后度量测试集中的查询数据和目标数据间的相似度从而得到跨模态匹配的结果。本发明可将不同模态的数据映射到统一空间度量,且在映射同时进行样本选择和特征学习,提高了匹配的鲁棒性和准确性。



1. 一种基于子空间的自步跨模态匹配方法,其特征在于,包括步骤:

S1,收集不同模态的数据样本建立跨模态数据库,并将所述跨模态数据库分为训练集和测试集;

S2,提取所述跨模态数据库中不同模态数据样本的特征向量;

S3,基于所述训练集中不同模态数据的特征向量建立目标函数并求解目标函数,得到与所述不同模态分别对应的映射矩阵;

S4,利用所述映射矩阵将所述测试集中数据样本的模态类别映射到同一空间;

S5,将映射到同一空间的测试集中同一模态类别的数据作为查询集,另一模态类别的数据作为目标集;

S6,通过度量数据间的相似度在所述目标集中查询与所述查询集中的一个数据样本相匹配的数据,从而得到跨模态匹配的结果。

2. 根据权利要求1所述方法,其特征在于,所述不同模态为双模态。

3. 根据权利要求2所述方法,其特征在于,所述双模态为文本和图像。

4. 根据权利要求3所述方法,其特征在于,对于文本模态数据样本提取隐狄克雷分布特征向量;对于图像多媒体数据样本提取尺度不变特征变换特征向量。

5. 根据权利要求1所述方法,其特征在于,步骤S3包括以下步骤:

S31,基于所述训练集中不同模态数据的特征向量建立目标函数;

S32,使用迭代算法求解所述目标函数,得到与所述不同模态分别对应的映射矩阵;

所述目标函数为:

$$\min_{\mathbf{U}_{\{a,b\}}, \mathbf{v}, \mathbf{Y}} \sum_{p \in \{a,b\}} \|\mathbf{U}_p^T \mathbf{X}_p - \mathbf{Y} \text{diag}(\mathbf{v})\|_F^2 + \alpha \sum_{p,q \in \{a,b\}} \text{Tr}(\mathbf{U}_p^T \mathbf{X}_p \mathbf{L}_{pq} \mathbf{X}_q^T \mathbf{U}_q) + \beta \sum_{p \in \{a,b\}} \|\mathbf{U}_p\|_F^2 - \frac{1}{k} \sum_i v_i$$

$$s.t. \mathbf{Y} \in \{0,1\}^{c \times n}, \sum_{i=1}^c Y_{i,j} = 1, \forall j \in [1, n].$$

其中,a,b分别代表两种模态,p,q的取值分别为a,b两种模态中的一种, $\mathbf{U}_p$ 是p模态对应的映射矩阵, $\mathbf{U}_p^T$ 是矩阵 $\mathbf{U}_p$ 的转置, $\mathbf{X}_p$ 是p模态数据特征, $\mathbf{U}_q$ 是q模态对应的映射矩阵, $\mathbf{X}_q$ 是q模态数据特征, $\mathbf{X}_q^T$ 是矩阵 $\mathbf{X}_q$ 的转置, $\mathbf{Y}$ 是聚类标记, $\mathbf{v} \in \mathbb{R}^n$ 代表损失项权重, $v_i$ 是向量 $\mathbf{v}$ 的第i个元素,k为常量,代表自步学习的学习率, $\alpha$ 和 $\beta$ 为权重参数, $\|\cdot\|_F$ 为求取F-范数操作, $\text{Tr}(\cdot)$ 是矩阵的迹, $\mathbf{L}_{pq}$ 是模态p和模态q的拉普拉斯矩阵,c,n分别为矩阵 $\mathbf{Y}$ 的行数和列数, $Y_{i,j}$ 是矩阵 $\mathbf{Y}$ 的第i行、第j列的元素。

6. 根据权利要求7所述方法,其特征在于,步骤S32包括步骤:

S321,使用k均值算法对文本特征聚类以初始化聚类标记 $\mathbf{Y}$ ,然后创建多模态图,多模态图是由不同模态数据特征之间的语义聚类标记和同一模态中数据特征之间的相似性构建;再计算多模态图约束拉普拉斯矩阵 $\mathbf{L}$ ,设置映射矩阵 $\mathbf{U}_p, p \in \{a, b\}$ 初始值都为单位矩阵,设置最大迭代次数N,迭代次数的初始值为1;

S322,计算每个样本的经验损失并与阈值k比较,得出每个样本的损失项权重 $v_i$ ;

$$v_i = \begin{cases} 1 & \text{if } \ell_i \leq \frac{1}{k} \\ 0 & \text{if } \ell_i > \frac{1}{k} \end{cases}$$

其中,  $\ell_i = \|(\mathbf{U}_p^T \mathbf{x}_p^i - \mathbf{y}^i)\|_F^2$  是第i个样本的损失函数;

S323, 通过求解下式中的线性问题求得映射矩阵  $\mathbf{U}_p$ :

$$\begin{aligned} & (\mathbf{X}_p \mathbf{V} \mathbf{V}^T \mathbf{X}_p^T + \alpha \mathbf{X}_p \mathbf{L}_{pp} \mathbf{X}_p^T + \beta \mathbf{I}) \mathbf{U}_p \\ & = \mathbf{X}_p \mathbf{V} \mathbf{V}^T \mathbf{Y}^T - \alpha \mathbf{X}_p \mathbf{L}_{pq} \mathbf{X}_q^T \mathbf{U}_q \end{aligned}$$

其中,  $\mathbf{Y} \in \{0, 1\}^{c \times n}$  为类别标签矩阵,  $n$  为特征向量的个数,  $c$  为模态类别的个数,  $\mathbf{I}$  为单位矩阵,  $\mathbf{V} = \text{diag}(\mathbf{v})$ ,  $\mathbf{v} \in \mathbb{R}^n$  代表损失项的权重,  $\mathbf{V}^T$  是矩阵  $\mathbf{V}$  的转置;

S324, 一列一列地优化聚类标记  $\mathbf{Y}$ , 即优化聚类标记  $\mathbf{Y}$  其中一列时, 固定其它所有列, 一次学习  $\mathbf{Y}$  的一列, 求解  $\mathbf{Y}$  的目标函数等价于:

$$\begin{aligned} & \min_{\mathbf{Y}} \text{Tr}(\mathbf{V}^T \mathbf{Y}^T \mathbf{Y} \mathbf{V}) + \alpha \text{Tr}(\mathbf{E} \mathbf{Y}^T \mathbf{Y} \mathbf{F}^T) \\ & \quad - \text{Tr}(\mathbf{G} \mathbf{Y}^T) - \text{Tr}(\mathbf{H} \mathbf{Y}^T) \\ & \text{s.t. } \mathbf{Y} \in \{0, 1\}^{c \times n}, \sum_i Y_{i,j} = 1, \forall j \in [1, n]. \end{aligned}$$

其中  $\mathbf{E} = \mathbf{U}_a^T \mathbf{X}_a$ ,  $\mathbf{F} = \mathbf{U}_b^T \mathbf{X}_b$ ,  $\mathbf{G} = \mathbf{U}_a^T \mathbf{X}_a \mathbf{V} \mathbf{V}^T$ ,  $\mathbf{H} = \mathbf{U}_b^T \mathbf{X}_b \mathbf{V} \mathbf{V}^T$ , 通过计算, 将上式转化为:

$$\begin{aligned} & \min_{\mathbf{y}} \mathbf{y}^T (2\mathbf{Y} \mathbf{V}^T \mathbf{v} + \alpha \mathbf{Y} \mathbf{F}^T \mathbf{e} + \alpha \mathbf{Y} \mathbf{E}^T \mathbf{f} - \mathbf{g} - \mathbf{h}) \\ & \text{s.t. } \mathbf{y} \in \{1, 0\}^{c \times 1}, \sum_i y_i = 1 \end{aligned}$$

其中,  $\mathbf{v}$  是矩阵  $\mathbf{V}$  的第  $i$  列,  $\mathbf{V}$  是矩阵  $\mathbf{V}$  移除第  $i$  列后得到的矩阵;  $\mathbf{e}$  是矩阵  $\mathbf{E}$  的第  $i$  列,  $\mathbf{E}$  是矩阵  $\mathbf{E}$  移除第  $i$  列后得到的矩阵;  $\mathbf{f}$  是矩阵  $\mathbf{F}$  的第  $i$  列,  $\mathbf{F}$  是矩阵  $\mathbf{F}$  移除第  $i$  列后得到的矩阵;  $\mathbf{g}$  是矩阵  $\mathbf{G}$  的第  $i$  列,  $\mathbf{G}$  是矩阵  $\mathbf{G}$  移除第  $i$  列后得到的矩阵;  $\mathbf{h}$  是矩阵  $\mathbf{H}$  的第  $i$  列,  $\mathbf{H}$  是矩阵  $\mathbf{H}$  移除第  $i$  列后得到的矩阵, 以上对  $\mathbf{y}$  的优化使用如下公式解决:

$$\mathbf{y}_i = \begin{cases} 1 & i = \mathbf{h}(\mathbf{m}) \\ 0 & \text{otherwise} \end{cases}$$

其中  $\mathbf{m} = 2\mathbf{Y} \mathbf{V}^T \mathbf{v} + \alpha \mathbf{Y} \mathbf{E}^T \mathbf{f} - \mathbf{g} - \mathbf{h}$ ,  $\mathbf{h}(\mathbf{m})$  返回  $\mathbf{m}$  中最大值的索引, 经过 2~3 内部迭代, 可以获得完整的  $\mathbf{Y}$ ;

S325,  $\mathbf{W}$  为多模态图的相似性矩阵, 通过下式计算  $\mathbf{W}$ :

$$\mathbf{W} = \begin{bmatrix} \gamma \mathbf{W}^a & \mathbf{W}^{ab} \\ \mathbf{W}^{ba} & \gamma \mathbf{W}^b \end{bmatrix}$$

其中,  $\gamma$  为模态间的相似性矩阵与同一模态内的局部相似性的权重参数, 取  $\gamma = 1$ ,  $\mathbf{W}^{ab} = \mathbf{W}^{ba} = \mathbf{Y}^T \mathbf{Y}$  是模态间的相似性矩阵,  $\mathbf{W}^a$  和  $\mathbf{W}^b$  为同一模态内的局部相似性, 采用高斯核函数

$d(\mathbf{x}_p^i, \mathbf{x}_p^j) = e^{-\frac{\|\mathbf{x}_p^i - \mathbf{x}_p^j\|^2}{2\sigma^2}}$  度量局部相似性:

$$W_{ij}^p = \begin{cases} d(\mathbf{x}_p^i, \mathbf{x}_p^j) & \text{if } \mathbf{x}_p^i \in N_r(\mathbf{x}_p^j) \text{ or } \mathbf{x}_p^j \in N_r(\mathbf{x}_p^i), \\ 0 & \text{otherwise} \end{cases}$$

其中  $N_r(\cdot)$  是最邻近的  $r$  个样本的集合;

S326,判断此时的迭代次数是否小于最大迭代次数N,若是则转向S322继续迭代;若否则停止迭代,得到所述映射矩阵 $U_a$ 和 $U_b$ 。

7.根据权利要求1所述方法,其特征在于,度量数据样本之间的相似度使用余弦距离来度量。

## 一种基于子空间的自步跨模态匹配方法

### 技术领域

[0001] 本发明涉及模式识别技术领域,尤其是一种基于子空间的自步跨模态匹配方法。

### 背景技术

[0002] 现实中的数据往往具有多种模态。比如网页数据既包含图片信息,又包含文本信息;视频数据同时包含音频信息和图片信息。跨模态匹配的根本任务是以一种模态作为查询条件,匹配出与之相似的异质模态信息。传统跨模态匹配方法大多是监督学习方法,它们都是通过语义标签来减少异质模态之间的鸿沟,但是无法处理无标签信息,人工标记数据又是一个费时费力的工作。此外,有些无监督的方法没有考虑特征的判定性和相关性和样本间的语义相似性,无法满足人们的日常需求。因此亟需一种高效的无监督跨模态匹配方法。

### 发明内容

[0003] 本发明目的在于解决上述技术问题而提供一种基于子空间的无监督自步跨模态匹配方法,能有效地跨越不同模态媒体之间的语义鸿沟,进而使得跨媒体搜索引擎返回的结果更加准确。

[0004] 本发明是这样实现的,一种基于子空间的自步跨模态匹配方法,包括步骤:

[0005] S1,收集不同模态的数据样本建立跨模态数据库,并将所述跨模态数据库分为训练集和测试集;

[0006] S2,提取所述跨模态数据库中不同模态数据样本的特征向量;

[0007] S3,基于所述训练集中不同模态数据的特征向量,得到与所述不同模态分别对应的映射矩阵;

[0008] S4,利用所述映射矩阵将所述测试集中数据样本的模态类别映射到同一空间;

[0009] S5,将映射到同一空间的测试集中同一模态类别的数据作为查询集,另一模态类别的数据作为目标集;

[0010] S6,通过度量数据间的相似度,在所述目标集中查询与所述查询集中的一个数据样本相匹配的数据,从而得到跨模态匹配的结果。

[0011] 根据本发明的方法,可以将不同模态的多媒体数据映射入统一空间进行度量,且在映射的同时进行了样本选择与特征学习,从而提高了检索的鲁棒性和准确性,具有良好的运用前景。

### 附图说明

[0012] 图1是本发明提供的基于子空间的自步跨模态匹配方法的流程图;

[0013] 图2是在图像和文本两个模态下子空间匹配的具体模型图。

### 具体实施方式

[0014] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明进一步详细说明。

[0015] 本发明通过学习两个映射矩阵,将不同模态的数据映射到同一个子空间中去,并在映射的同时进行样本选择和特征学习,并使用多模态图约束保持数据模态内和模态间的相似性;在学习到的子空间中度量不同模态的数据的相似性,以此来实现跨模态匹配。

[0016] 参见图1所示,一种基于子空间的自步跨模态匹配方法,包括以下步骤:

[0017] 步骤S1,收集不同模态的数据样本,建立跨模态数据库,并将所述跨模态数据库分为训练集和测试集;

[0018] 需要说明的是,本发明中所述不同模态可以为文本、图像等模态。

[0019] 为了便于说明,下面以文本和图像两个模态的数据为例进行解释本发明。

[0020] 具体的,将所述跨模态数据库分为训练集和测试集时,可根据实际需要进行划分,如可将所述跨模态数据库中的80%的数据划分为训练集,剩下的20%的数据划分为测试集,形成训练样本库与测试样本库。

[0021] 步骤S2,提取所述跨模态数据库中不同模态数据样本的特征向量,形成相应的不同模态数据特征库;

[0022] 本步骤中,不但要提取训练集的训练样本库中不同模态数据样本的特征向量,也要提取测试集的测试样本库中不同模态数据样本的特征向量,从而形成训练集与测试集的文本特征库与图片特征库;

[0023] 本发明中,对于文本数据库中的文本模态数据样本和图像数据库的图像模态数据样本,可分别使用LDA算法与SIFT算法进行特征提取;

[0024] LDA算法即隐狄克雷分布(Latent Dirichlet Allocation,LDA)算法,SIFT算法即尺度不变特征变换(Scale-Invariant Feature Transform,SIFT)算法。

[0025] 步骤S3,基于所述训练集中提取的不同模态数据的特征向量,通过子空间自步学习(即通过建立目标函数并通过迭代算法求解目标函数),得到与所述不同模态分别对应的特征映射矩阵;

[0026] 对步骤S3详细介绍前,先介绍矩阵运算符:对于矩阵 $\mathbf{M} \in \mathbb{R}^{n \times m}$ ,第 $i$ 行和第 $j$ 列分别为 $\mathbf{m}_i$ 和 $\mathbf{m}^j$ , $M_{i,j}$ 是矩阵的第 $i$ 行和第 $j$ 列元素。矩阵 $\mathbf{M}$ 的Frobenius范数为 $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \|\mathbf{m}_i\|_2^2}$ ,矩阵 $\mathbf{M}$ 的迹为 $\text{Tr}(\mathbf{M}) = \sum_i M_{i,i}$ 。

[0027] 假设有图片和文本两个特征集, $\mathbf{X}_a = [\mathbf{x}_1^a, \mathbf{x}_2^a, \dots, \mathbf{x}_n^a] \in \mathbb{R}^{d_a \times n}$ , $\mathbf{X}_b = [\mathbf{x}_1^b, \mathbf{x}_2^b, \dots, \mathbf{x}_n^b] \in \mathbb{R}^{d_b \times n}$ ,其中 $d_i$ 是第 $i$ 个模态的维度, $n$ 是训练图像-文本对的数目。每一对图像-文本对 $\{\mathbf{x}_i^a, \mathbf{x}_i^b\}$ 具有相同的隐含内容,且属于相同类别,即硬配对约束,但每个图像-文本对的离散标签是未知的。

[0028] 所述步骤S3包括以下步骤:

[0029] 步骤S31,基于所述训练集中不同模态数据的特征向量建立目标函数:

$$[0030] \quad \min_{\mathbf{U}_{\{a,b\}}, \mathbf{v}, \mathbf{Y}} \sum_{p \in \{a,b\}} \|(\mathbf{U}_p^T \mathbf{X}_p - \mathbf{Y}) \text{diag}(\mathbf{v})\|_F^2 + \alpha \sum_{p,q \in \{a,b\}} \text{Tr}(\mathbf{U}_p^T \mathbf{X}_p \mathbf{L}_{pq} \mathbf{X}_q^T \mathbf{U}_q) + \beta \sum_{p \in \{a,b\}} \|\mathbf{U}_p\|_F^2 - \frac{1}{k} \sum_i v_i$$

$$[0031] \quad \text{s.t. } \mathbf{Y} \in \{0,1\}^{c \times n}, \sum_{i=1}^c Y_{i,j} = 1, \forall j \in [1,n].$$

[0032] 其中, $\alpha$ 和 $\beta$ 为权重参数,取值范围为 $10^{-6} \sim 10^2$ ;

[0033] 其中, $a, b$ 分别代表两种模态, $p, q$ 的取值分别为 $a, b$ 两种模态中的一种, $U_p$ 是 $p$ 模态对应的映射矩阵, $U_p^T$ 是矩阵 $U_p$ 的转置, $X_p$ 是 $p$ 模态数据特征, $U_q$ 是 $q$ 模态对应的映射矩阵, $X_q$ 是 $q$ 模态数据特征, $X_q^T$ 是矩阵 $X_q$ 的转置, $Y$ 是聚类标记, $\mathbf{v} \in \mathbb{R}^n$ 代表损失项权重, $v_i$ 是向量 $\mathbf{v}$ 的第 $i$ 个元素, $k$ 为常量,代表自步学习的学习率, $\alpha$ 和 $\beta$ 为权重参数, $\|\cdot\|_F$ 为求取 $F$ -范数操作, $\text{Tr}(\cdot)$ 是矩阵的迹, $L_{pq}$ 是模态 $p$ 和模态 $q$ 的拉普拉斯矩阵, $c, n$ 分别为矩阵 $Y$ 的行数和列数, $Y_{i,j}$ 是矩阵 $Y$ 的第 $i$ 行、第 $j$ 列的元素。

[0034] 步骤S32,求解所述目标函数(如通过迭代算法)得到与所述不同模态分别对应的特征映射矩阵:

[0035] 所述步骤S32进一步包括以下步骤:

[0036] 步骤S321,使用 $k$ 均值算法对文本特征聚类以初始化聚类标记 $Y$ ,然后计算多模态图约束拉普拉斯矩阵 $L$ ,设置映射矩阵 $U_p, p \in \{a, b\}$ 的初始值为单位矩阵,设置最大迭代次数 $N$ ,迭代次数的初始值为1;

[0037] 使用 $k$ 均值算法对文本特征聚类以初始化聚类标记 $Y$ ,是指先对文本特征聚类,然后将文本特征聚类结果作为聚类标记 $Y$ 的初始值;

[0038] 多模态图由两部分组成:一是不同模态样本特征之间的相似性;二是同一模态内不同样本特征之间的相似性。前者相似性的计算是根据聚类标签 $Y$ 得到;后者是通过计算高斯核函数得到。

[0039] 步骤S322,计算每个样本的经验损失,将其与阈值 $k$ 比较,得出损失项权重 $v_i$ ;

$$[0040] \quad v_i = \begin{cases} 1 & \text{if } \ell_i \leq \frac{1}{k}, \\ 0 & \text{if } \ell_i \geq \frac{1}{k}, \end{cases}$$

[0041] 其中, $\ell_i = \|(\mathbf{U}_p^T \mathbf{x}_p^i - \mathbf{y}^i)\|_F^2$ 是第 $i$ 个样本的损失函数。

[0042] 步骤S323,通过求解下式中的线性问题来求得映射矩阵 $U_p$ :

$$[0043] \quad \begin{aligned} & (\mathbf{X}_p \mathbf{V} \mathbf{V}^T \mathbf{X}_p^T + \alpha \mathbf{X}_p \mathbf{L}_{pp} \mathbf{X}_p^T + \beta \mathbf{I}) \mathbf{U}_p \\ & = \mathbf{X}_p \mathbf{V} \mathbf{V}^T \mathbf{Y}^T - \alpha \mathbf{X}_p \mathbf{L}_{pq} \mathbf{X}_q^T \mathbf{U}_q \end{aligned}$$

[0044] 其中, $Y \in \{0, 1\}^{c \times n}$ 为类别标签矩阵, $n$ 为特征向量的个数, $c$ 为模态类别的个数, $I$ 为单位矩阵, $V = \text{diag}(\mathbf{v}), \mathbf{v} \in \mathbb{R}^n$ 代表损失项的权重, $V^T$ 是矩阵 $V$ 的转置;

[0045] 步骤S324,一列一列地优化 $Y$ ,即优化 $Y$ 的其中一列时,固定其它所有列,一次学习 $Y$ 的一列。求解 $Y$ 的目标函数等价于:

$$[0046] \quad \begin{aligned} & \min_{\mathbf{Y}} \text{Tr}(\mathbf{V}^T \mathbf{Y}^T \mathbf{Y} \mathbf{V}) + \alpha \text{Tr}(\mathbf{E} \mathbf{Y}^T \mathbf{Y} \mathbf{F}^T) \\ & - \text{Tr}(\mathbf{G} \mathbf{Y}^T) - \text{Tr}(\mathbf{H} \mathbf{Y}^T), \\ & \text{s.t. } \mathbf{Y} \in \{0, 1\}^{c \times n}, \sum_i Y_{i,j} = 1, \forall j \in [1, n]. \end{aligned}$$

[0047] 其中 $\mathbf{E} = \mathbf{U}_a^T \mathbf{X}_a, \mathbf{F} = \mathbf{U}_b^T \mathbf{X}_b, \mathbf{G} = \mathbf{U}_a^T \mathbf{X}_a \mathbf{V} \mathbf{V}^T, \mathbf{H} = \mathbf{U}_b^T \mathbf{X}_b \mathbf{V} \mathbf{V}^T$ 。通过计算,将上式转化为:

$$\begin{aligned}
 & \min_{\mathbf{y}} \mathbf{y}^T (2\mathbf{YV}^T \mathbf{v} + \alpha \mathbf{YF}^T \mathbf{e} + \alpha \mathbf{YE}^T \mathbf{f} - \mathbf{g} - \mathbf{h}) \\
 [0048] \quad & \text{s.t. } \mathbf{y} \in \{1, 0\}^{\text{col}}, \sum_i y_i = 1
 \end{aligned}$$

[0049] 其中,  $\mathbf{v}$  是矩阵  $V$  的第  $i$  列,  $V$  是矩阵  $V$  移出第  $i$  列后得到的矩阵;  $\mathbf{e}$  是矩阵  $E$  的第  $i$  列,  $E$  是矩阵  $E$  移出第  $i$  列后得到的矩阵;  $\mathbf{f}$  是矩阵  $F$  的第  $i$  列,  $F$  是矩阵  $F$  移出第  $i$  列后得到的矩阵;  $\mathbf{g}$  是矩阵  $G$  的第  $i$  列,  $G$  是矩阵  $G$  移出第  $i$  列后得到的矩阵;  $\mathbf{h}$  是矩阵  $H$  的第  $i$  列,  $H$  是矩阵  $H$  移出第  $i$  列后得到的矩阵。以上对  $\mathbf{y}$  的优化使用如下公式解决:

$$[0050] \quad \mathbf{y}_i = \begin{cases} 1 & \mathbf{i} = \mathbf{h}(\mathbf{m}) \\ 0 & \text{otherwise} \end{cases}$$

[0051] 其中  $\mathbf{m} = 2\mathbf{YV}^T \mathbf{v} + \alpha \mathbf{YE}^T \mathbf{e} + \alpha \mathbf{YF}^T \mathbf{f} - \mathbf{g} - \mathbf{h}$ ,  $\mathbf{h}(\mathbf{m})$  返回  $\mathbf{m}$  中最大值的索引。经过 2~3 内部迭代, 可以获得完整的  $\mathbf{Y}$ 。

[0052] 步骤 S325,  $\mathbf{W}$  为多模态图的相似性矩阵, 通过下式计算  $\mathbf{W}$ :

$$[0053] \quad \mathbf{W} = \begin{bmatrix} \gamma W^a & W^{ab} \\ W^{ba} & \gamma W^b \end{bmatrix},$$

[0054] 其中,  $\gamma$  为模态间的相似性矩阵与同一模态内的局部相似性的权重参数, 取  $\gamma = 1$ ,  $W^{ab} = W^{ba} = \mathbf{Y}^T \mathbf{Y}$  是模态间的相似性矩阵,  $W^a$  和  $W^b$  为同一模态内的局部相似性, 采用高斯核函

数  $d(\mathbf{x}_p^i, \mathbf{x}_p^j) = e^{-\frac{\|\mathbf{x}_p^i - \mathbf{x}_p^j\|^2}{2\sigma^2}}$  度量局部相似性:

$$[0055] \quad W_{ij}^p = \begin{cases} d(\mathbf{x}_p^i, \mathbf{x}_p^j) & \text{if } \mathbf{x}_p^i \in N_r(\mathbf{x}_p^j) \text{ or } \mathbf{x}_p^j \in N_r(\mathbf{x}_p^i), \\ 0 & \text{otherwise} \end{cases}$$

[0056] 其中  $N_r(\cdot)$  是最邻近的  $r$  个样本的集合。

[0057] 步骤 S326, 判断此时的迭代次数是否小于最大迭代次数  $N$ , 若是则转向步骤 S322 继续迭代; 若否则停止迭代, 得到所述映射矩阵  $U_a$  和  $U_b$ 。

[0058] 步骤 S4, 将所述测试集中多模态数据样本的模态类别, 利用步骤 S3 得到的特征映射矩阵映射到同一子空间;

[0059] 步骤 S5, 将映射到同一子空间后的测试集中同一模态类别的数据作为查询集, 另一模态类别的数据作为目标集;

[0060] 步骤 S6, 对于所述查询集中的一个数据样本, 通过度量数据间的相似度, 得到所述目标集中与查询集中的一个数据样本相匹配的数据, 从而得到跨模态匹配结果。

[0061] 本发明中, 所述数据之间的相似度可以使用余弦距离来度量, 对于余弦距离来说, 距离越近说明这两个数据越相似。

[0062] 附图 2 中, 图片提取 SIFT 特征后形成相应的方框图, 文本提取 LDA 特征后形成相应的柱状图, 在图片提取 SIFT 特征之后形成的特征空间  $\mathfrak{R}^i$  中, 方块代表每个图片的 SIFT 特征在图像特征空间下的点, 虚线圈表示圈中的这些方块点属于同一个聚类标记。文本提取 LDA 特征之后的文本特征空间  $\mathfrak{R}^t$  中, 虚线圈中的圆圈代表每个文本的 LDA 特征在特征空间下的点。虚线圈是表示圈中的这些点属于同一个聚类标记, 语义组带箭头虚线表示指向的图像特征空间与文本特征空间中分别属于同一个聚类标记的对应的特征点对应具有相同的聚



类标签,配对带箭头实线表示分别属于同一个聚类标记的对应形成的图像-文本数据对,在图片特征空间和文本特征空间投影后的公共子空间 $\mathfrak{R}^c$ 中,通过自步学习选出来训练的样本对,形成自步学习对。

[0063] 为了验证本发明方法的有效性,接下来将本发明方法应用于Wiki多模态数据库。该数据库中包含10个语义范畴的多模态数据,包括2866对相同语义信息的文本和图像,随机选取其中1300对文本和图像样本作为训练集,选取其中1566对文本和图像样本作为测试集,应用本发明方法的步骤如下:参见图2所示;

[0064] 1)对于所述数据库中的数据,根据模态类别的不同提取不同的特征向量:对于文本模态提取10维LDA语义特征,对于图像模态提取128维SIFT特征;

[0065] 2)将所述训练集中文本模态特征矩阵和图像模态特征矩阵,构造目标函数,通过最小化目标函数学习得到文本和图像对应的两个映射矩阵,所述映射矩阵的学习过程通过迭代过程实现;

[0066] 3)根据所述模态类别的不同,将测试集中的文本数据的特征向量和图像数据的特征向量通过学习得到的映射矩阵映射到同一子空间中;

[0067] 4)将测试集中的文本数据作为查询集,图像数据作为目标集,在查询集中给定一个文本文档,能够得到目标集中和文本文档最相似(即余弦距离最近)的图像。

[0068] 综上,本发明提出的基于子空间的自步跨模态匹配方法,能有效地跨越了不同模态数据之间的语义鸿沟,与传统跨模态方法比较,易于实现、性能稳定,识别准确率高。另外,本方法用自步学习框架,即按简单到复杂的顺序选择样本进行学习(简单和复杂的定义根据每个样本的经验损失确定),避免目标函数陷入局部极小值,提高聚类精度;同时使用多模态图约束,保持数据映射前后结构的相似性,挖掘出一个判别的隐空间,从而实现跨模态匹配。

[0069] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

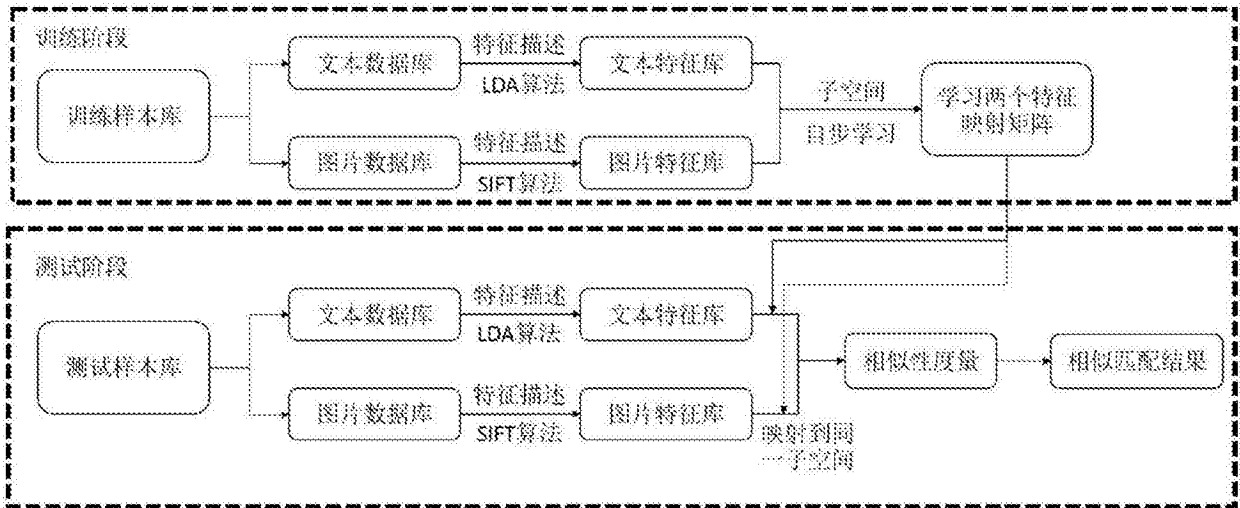


图1

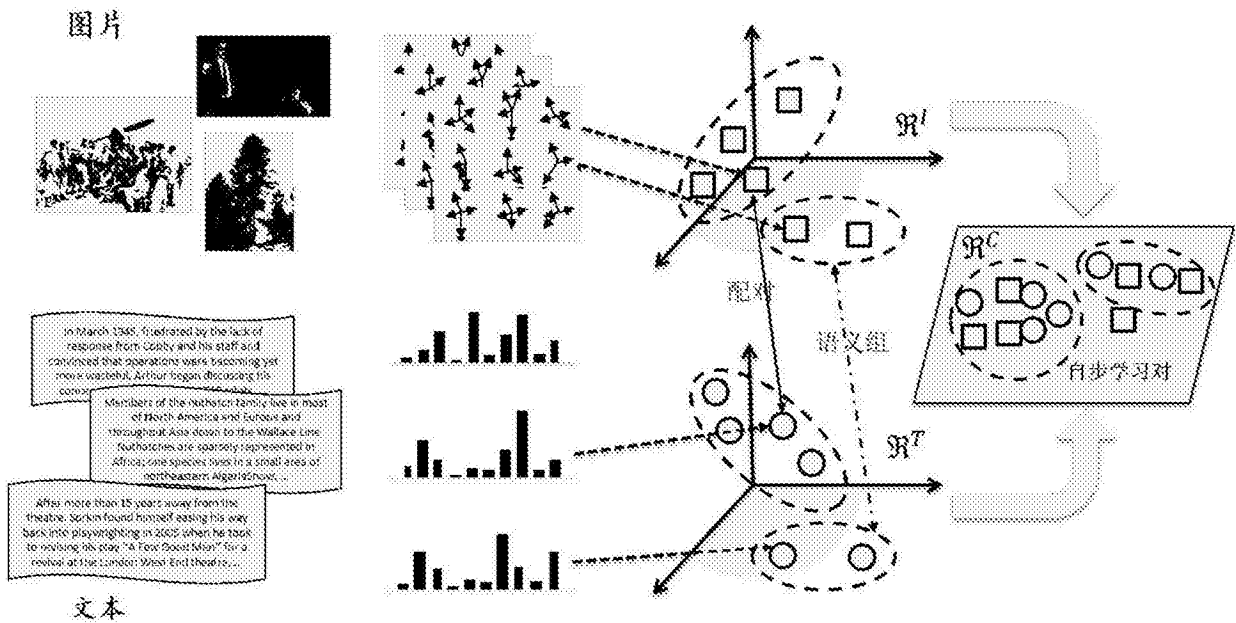


图2