doi:10.11887/j.cn.201801014 http://journal.nudt.edu.cn

基于局部敏感哈希算法和神经网络学习的跨媒体检索方法*

白 亮,贾玉华,王昊冉,谢毓湘,于天元
(国防科技大学 系统工程学院,湖南 长沙 410073)

摘 要:为了提高跨媒体检索的效率,可行的方法是降低数据集中不相关内容的比例。采用局部敏感哈希算法将图像数据映射到汉明空间并利用神经网络学习将文本数据映射到汉明空间的哈希函数,提出一种可以显著提高数据集中相关文件比例的高效跨媒体检索方法。实验结果表明,提出的方法能够有效去除数据集中的不相关内容,相对于已有的跨媒体检索方法,其有效提高了检索效率与准确率。

关键词:跨媒体检索;神经网络算法;局部敏感哈希算法;多模态数据索引
中图分类号:TP391 **文献标志码**:A **文章编号**:1001-2486(2018)01-093-06

Cross-media retrieval based on locality-sensitive hashing and neural network algorithms

BAI Liang, *JIA Yuhua*, *WANG Haoran*, *XIE Yuxiang*, *YU Tianyuan*
(College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: To efficiently retrieve in multimodal data, it is essential to reduce the proportion of irrelevant documents. The image data were projected to the Hamming space by using the locality-sensitive hashing algorithm, the text data were mapped on the hashing function of Hamming space by employing the neural network learning, and then a novel cross-media retrieval approach was proposed to reduce the proportion of irrelevant documents. The experiment shows that the proportion of the relevant documents can be much improved in the proposed method. Assessments on the two public datasets also demonstrate the efficacy and the accuracy of the proposed retrieval method when compared to the baselines.

Key words: cross-media retrieval; neural network algorithm; locality-sensitive hashing algorithm; multimodal data indexing

在跨媒体大数据时代,无时无刻不在产生的海量多模态信息带来了巨大的跨媒体检索需求。例如:用文本来搜索图像或视频;反之亦然,如维基百科上的一个词条通常包含文本描述和示例图像。这些信息的检索需要构建跨媒体索引与学习方法。与传统的单一媒体检索相比,跨媒体检索的核心问题是如何挖掘不同媒体表示的相同或相关语义对象之间的关联。

针对这一具有挑战性的问题,已有研究提出了众多的解决方法。已有的跨媒体检索方法主要分为两类。一类是基于主题的方法:主要是通过主题比例分析对不同模态的数据之间的相关性进行建模[1]。文献[2]通过 CORR-LDA(correspondence-Latent Dirichlet allocation)挖掘图像与文本标注之间在主题层次的关系;文献[3]将马尔可夫随机域与传统 LDA 方法结合,提出了用简短的文字检索图像的有向和无向概率图的组合的多模态文档随机域(Multi-modal Document Random Field, MDRF)模型;文献[4]提出一种利用多个媒体类型的微博信息,构建社会事件的可视化总结的多媒体社会事件自动摘要框架。另一类是基于子空间的方法:核心是寻求使不同模态数据相关性最大化的子空间[5]。Ranjan 等针对具有多标注信息的多模态数据提出了一种可以将多标注形式的高层语义信息用于子空间学习的 ml-CCA 方法[6];文献[7]提出的 T-V CCA 模型中引入语义的观点,以提高子空间中不同类别的多模态数据的分类准确率;文献[8]提出了一种 Bi-CMSRM 方法,从优化双向列表排序问题的角度构建了适用于跨媒体检索的计算模型。近年来,提出了很多利用深度学习求取跨媒体数据共享子空间的方法。Yan 等将全连接层叠加在一起来提取文本与图像的特征,并且使用深度典型相关分析(Depth Canonical Correlation Analysis, DCCA)[9]学习图像和文本的共享子空间,进而完成图像语文本的匹配;Yu 等将深度语义匹配模型和 AlexNet 结合

* 收稿日期:2016-11-03
基金项目:国家自然科学基金资助项目(61571453);湖南省自然科学基金资助项目(14JJ3010)
作者简介:白亮(1978—),男,陕西清涧人,副教授,博士,硕士生导师,E-mail:xabpz@163.com

起来提出了一种用于双向文本图像检索的深度双流网络[10]。

现有方法存在一个共同问题,即仅仅只考虑了跨媒体检索方法本身而忽略了对文档集的一些可行的优化处理,由于文档集中存在大量与查询不相关的文档,因此在进行精确查询之前对文档集进行预处理,提高文档集中相关文件所占比例,对提高检索效率来说具有重要意义。目前,局部敏感哈希(Locality Sensitive Hashing, LSH)算法[11]已广泛应用于图像或文本检索并表现出很好的性能[12]。本文提出一种基于 LSH 算法与神经网络学习的快速跨媒体检索(Fast Cross-Media Retrieval,FCMR)方法,通过消除大量与查询无关的文档内容并获得一组查询的最近邻,最终在查询文档的最近邻范围内更高效地进行检索任务。

1 FCMR 模型综述与框架描述

为了使符号和算法表述更加简洁,下面以文本和图像两个模态为例描述提出的模型,模型可以很容易地扩展到其他模态。提出的 FCMR 的框架如图 1 所示。FCMR 包括局部敏感哈希和哈希函数学习两个阶段。在局部敏感哈希阶段,采用局部敏感哈希算法将图像数据映射到哈希桶;在哈希函数学习阶段,采用神经网络学习将文本数据映射到哈希桶的哈希函数 Ht。FCMR 的过程是首先通过局部敏感哈希算法将图像数据映射到 m 个哈希表 $\boldsymbol{G}=[g_1,g_2,\cdots,g_m]\in\mathbf{R}^{k\times m}$ 的哈希桶内,其中 $\boldsymbol{G}$ 为 m 个哈希表的集合,g_j 表示第 j 个哈希表,k 是哈希桶对应哈希码的长度;其次通过神经网络学习将文本数据分别映射到 m 个哈希表内其对应的哈希桶内的哈希函数 $Ht=(Ht^{(1)},Ht^{(2)},\cdots,Ht^{(m)})$,$Ht^{(j)}(1\leqslant j\leqslant m)$ 表示学习到的对应于第 j 个哈希表的哈希函数。

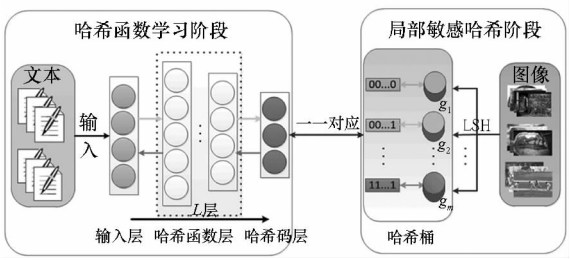


图 1 FCMR 方法框架图

Fig. 1 Flowchart of FCMR

设 $\boldsymbol{T}=[\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_{nt}]\in\mathbf{R}^{dt\times nt}$,其中 $\boldsymbol{T}$ 是文本数据的矩阵描述,相应地,$\boldsymbol{P}=[\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_{np}]\in\mathbf{R}^{dp\times np}$,其中 $\boldsymbol{P}$ 是图像数据的矩阵描述。其中,dt 和 dp 分别代表文本和图像特征的维度,$\boldsymbol{t}_i$ 与 $\boldsymbol{p}_i$ 一一对应,即 $nt=np$,在下面的内容里用 n 代替 nt 和 np。

如果用局部敏感哈希算法获得了 m 个哈希表,那么需要设计 m 个与其对应的神经网络将文本数据映射到 m 个哈希表中它们对应的哈希桶内。基于神经网络学习到的哈希函数与局部敏感哈希阶段使用的局部敏感哈希函数,可以对多模态数据建立索引从而进行高效的跨媒体检索任务。给定一个查询文本,$Ht^{(j)}$ 将其映射到 m 个哈希表中的哈希桶内,则这些哈希桶内存储的图像文件就组成了该查询文本的最近邻,进而就可在这个最近邻范围内进行精确地检索;给定一个查询图像,通过局部敏感哈希函数将其映射到 m 个哈希表中的哈希桶内,则这些哈希桶内存储的文本文件就组成了该查询图像的最近邻,进而就可在这个邻居列表里进行精确检索。

2 局部敏感哈希算法

局部敏感哈希算法主要用来解决高维空间中点的近似最近邻搜索问题[13]。局部敏感哈希函数定义为:

$$h_{\boldsymbol{r}}(\boldsymbol{p}_i)=\begin{cases}1 & \boldsymbol{r}^{\mathrm{T}}\boldsymbol{p}_i\geqslant 0\\ 0 & \text{else}\end{cases}\tag{1}$$

其中,超平面向量 $\boldsymbol{r}$ 符合 multi Gaussian N(0,1)分布,文献[13]中对式(1)进行了证明,这里不再赘述。

定义一系列哈希函数 $h_1,h_2,\cdots,h_n$,随机选取其中的 k 个函数组成函数 $g(x)$,不妨设选的是 h_1 到 h_k,则 $g(x)=(h_1(x),h_2(x),\cdots,h_k(x))$,由此选取 m 个 $g(x)$ 函数 $g_1(x),g_2(x),\cdots,g_m(x)$,每个函数对应一个哈希表。通过 m 个 $g(x)$ 函数将图像空间中的每一个样本 $\boldsymbol{p}_i$ 分别映射到 m 个哈希表中,这样每个图像样本都会在 m 个哈希表的某个哈希桶中出现。

那么 $\boldsymbol{p}_i$ 在第 j 个哈希表里对应的哈希桶可以表示为:

$$g_j(\boldsymbol{p}_i)=\langle h_1(\boldsymbol{p}_i),h_2(\boldsymbol{p}_i),\cdots,h_k(\boldsymbol{p}_i)\rangle,\quad 0<j\leqslant m,0<i\leqslant n\tag{2}$$

查询时,给定查询文本,利用 m 个 $g(x)$ 函数同时对查询文本进行映射,将与查询文本落在相同的哈希桶中的图像样本作为候选结果集,计算查询文本与候选结果集中的图像之间的距离并进行精确的检索排名。

通过局部敏感哈希算法,图像空间的样本 $\boldsymbol{p}_i$ $(0<i\leqslant n)$ 被映射到 m 个哈希表中,且每个 $\boldsymbol{p}_i$ $(0<i\leqslant n)$ 都会和与其相似的样本一起出现在 m

个哈希表的某个哈希桶。这样,每一个图像样本 $\boldsymbol{p}_i$ 都与第 $j(0<j\leqslant m)$ 个哈希表的某个哈希桶建立了联系。同时上面提到过,在模型中由于 $\boldsymbol{p}_i$ 和 $\boldsymbol{t}_i$ 是同一语义不同模态的描述,图像样本与文本样本是一一对应的,因此,每一个文本样本 $\boldsymbol{t}_i$ 也与第 $j(0<j\leqslant m)$ 个哈希表的某个哈希桶建立了联系。至此,得到了用于训练神经网络学习将文本样本 $\boldsymbol{t}_i$ 映射到第 $j(0<j\leqslant m)$ 个哈希表中其对应哈希桶的函数的训练样本:$(\boldsymbol{t}_i, \boldsymbol{Y}_i^{(j)})$ $(i\in 1,2,\cdots,n, j\in 1,2,\cdots,m)$,其中 $\boldsymbol{Y}_i^{(j)}$ 表示 $\boldsymbol{p}_i$ 对应于第 $j(0<j\leqslant m)$ 个哈希表中的哈希桶的哈希码。

3　神经网络设计

图 1 给出了哈希函数学习阶段神经网络结构,模型中使用到的 m 个神经网络 $NN^{(j)}$ $(j\in 1,2,\cdots,m)$ 具有相同的结构。每一个 $NN^{(j)}$ 有 L 层,其中输入层有 dt 个神经元对应于文本特征的维度,输出层有 k 个神经元对应于哈希码的 k 位,剩余的 $L-2$ 层用于学习哈希函数。将每一个 $\boldsymbol{t}_i\in \boldsymbol{T}$ 作为 $NN^{(j)}$ 的输入,可以得到神经网络各个层的输出 $\boldsymbol{t}_i^{(1)},\cdots,\boldsymbol{t}_i^{(L)}$。第 $l+1$ 层以 $\boldsymbol{t}_i^{(l)}$ 为输入,输出 $\boldsymbol{t}_i^{(l+1)}$:

$$\boldsymbol{t}_i^{(l+1)}=f^{(l+1)}(\boldsymbol{W}^{(l+1)}\boldsymbol{t}_i^{(l)}) \tag{3}$$

式中,$\boldsymbol{t}_i^{(l)}\in \mathbf{R}^k$ 和 $\boldsymbol{t}_i^{(l+1)}\in \mathbf{R}^k$ 分别为第 l 层和第 $l+1$ 层的特征表达;$\boldsymbol{W}^{(l+1)}$ 是转换矩阵;$f^{(l+1)}(\cdot)$ 是激活函数。神经网络学习到的哈希函数 $Ht^{(j)}$ 以 $\boldsymbol{t}_i$ 为输入并输出长度为 k 的哈希码:

$$Ht^{(j)}(\boldsymbol{t}_i)=\mathrm{sign}(\boldsymbol{t}_i^{(L-1)}) \tag{4}$$

式中,$\boldsymbol{t}_i^{(L-1)}\in \mathbf{R}^k$ 是一个 k 维实值向量,使用符号函数将 $\boldsymbol{t}_i^{(L-1)}$ 转化为哈希码。

由于符号函数受不可微性制约难以优化,因此在用神经网络学习哈希函数的阶段除去了符号函数,再在测试阶段重新加上[14-15]。

对于训练样本 $(\boldsymbol{t}_i,\boldsymbol{Y}_i^{(j)})$,$Ht^{(j)}(\boldsymbol{t}_i)$ 与 $\boldsymbol{Y}_i^{(j)}$ 应该相同,也就是 $\boldsymbol{t}_i^{(L-1)}$ 和 $\boldsymbol{Y}_i^{(j)}$ 应尽可能相等,基于最小方差定义损失函数:

$$SE(\boldsymbol{t}_i^{(L-1)},\boldsymbol{Y}_i^{(j)})=\frac{1}{2}\|\boldsymbol{t}_i^{(L-1)}-\boldsymbol{Y}_i^{(j)}\|_F^2 \tag{5}$$

式中,$\boldsymbol{t}_i^{(L-1)}$ 是未加符号函数神经网络对 $\boldsymbol{t}_i$ 的预测值。

4　FCMR 算法

FCMR 模型的训练过程包括局部敏感哈希阶段和神经网络哈希函数学习阶段。从局部敏感哈希阶段得到训练神经网络所需的训练样本 $(\boldsymbol{t}_i,\boldsymbol{Y}_i^{(j)})$,$i\in 1,2,\cdots,n, j\in 1,2,\cdots,m$。通过训练神经网络 $NN^{(j)}$ 可以使其学习到将 $\boldsymbol{t}_i$ 映射到 $\boldsymbol{Y}_i^{(j)}$ 的哈希函数。神经网络的训练分为预训练和参数调整。预训练可以更好地初始化网络参数并防止网络陷入局部最优解。将栈自编码器(Stacked AutoEncoder,SAE)应用于 FCMR 模型来顺序地训练 $NN^{(j)}$ 中的每一层以初始化网络参数。然后基于损失函数式(5),通过反向传播(Back Propagation,BP)算法来训练网络调整网络参数。最后基于所有文本样本的方差和 SSE(sum of standard error)设计了整体的损失函数如式(6)所示:

$$SSE(\boldsymbol{t}_i^{(L)},\boldsymbol{Y}_i^{(j)})=\frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{t}_i^{(L)}-\boldsymbol{Y}_i^{(j)}\|_F^2 \tag{6}$$

神经网络 $NN^{(j)}$ 学习到的函数 $Ht^{(j)}$ 应该很好地将文本样本数据映射到 j 个哈希表中其对应的哈希桶内。采用传统的反向传播算法来训练神经网络 $NN^{(j)}$,在测试阶段最终通过式(4)获得最终的哈希函数 $Ht^{(j)}$。FCMR 的算法过程见算法 1。

算法 1　FCMR 算法

Alg. 1　FCMR algorithm

Input:　data set $\boldsymbol{T},\boldsymbol{P},k,m$

Output:　The hash functions Ht and LSH hash functions

```
for j = 1 to m do
    (t, Y^(j)) ← LSH(P)
    NN^(j) ← SAE(T)
    repeat
        Pick a random training sample (t_i, Y_i^(j))
        Make a gradient step for SE
            SE(t_i^(L-1), Y_i^(j)) = 1/2 ||t_i^(L-1) − Y_i^(j)||_F^2
        Update NN^(j)
    until stopping criteria is met
end for
```

5　实验结果与分析

5.1　实验设置

1)数据集:在实验中选取了 Wiki① 和 NUS-WIDE② 两个数据集,每一个数据集都包含一一对应的文本和图像。两个数据集的详细信息见表 1。Wiki 和 NUS-WIDE 数据集图像和标签分

① http://www.svcl.ucsd.edu/projects/crossmodal/

② http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm

别为10类和81类。

表1　数据集详细信息

Tab. 1　Details of the data sets

数据集信息	Wiki	NUS-WIDE
BoVW 词典规模	1000	500
BoW 词典规模	5000	1000
平均标注数	117.5	7.73
训练集规模	1000	100 000
验证集规模①	866/866	5000/20 000
测试集规模①	1000/1000	5000/149 648

注:①分割顺序为查询/数据库,其中查询随机地从数据库中选取。

2)评价指标:为了评价提出的检索方法,使用以下四个评价指标进行结果评价:

召回率:邻居列表中与检索文档相关的文档占总文件集中相关文件的比例。

选择度[13]:邻居列表内的文档总数占总文档集的比例。

相关文件比例:邻居列表中与检索文档相关的文档占邻居列表内文档总数的比例。

MAP:平均精度均值。

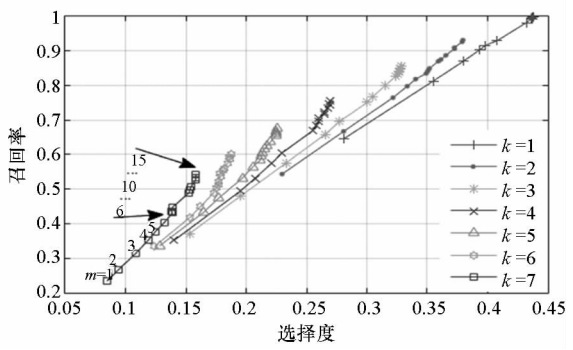
3)参数设置及调整:在实验中,所有神经网络设置为3层网络,其中输出层的激活函数为softmax,其余层的激活函数为sigmoid函数。虽然网络只有3层,但已经达到了满意的效果。在对两个数据集上各个类别中随机成比例采样的验证集上进行实验来选取FCMR模型的参数k和m,其中k为哈希码的长度,m为在局部敏感哈希阶段生成的哈希表个数。在Wiki数据集上,选择的k和m的取值范围分别为$\{1,2,\cdots,7\}$和$\{1,2,\cdots,15\}$;在NUS-WIDE数据集上,选择的k和m的取值范围分别为$\{2,3,\cdots,9\}$和$\{1,2,\cdots,20\}$。最终选择表现最好的参数对作为后续实验的参数。

5.2　结果及分析

从文本检索图像(T2I)和图像检索文本(I2T)两个方面进行了实验。

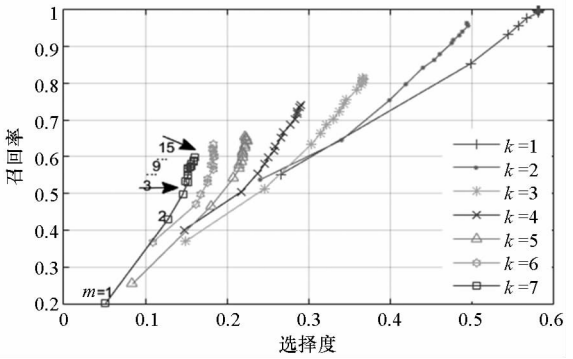
图2给出了从文本检索图像和图像检索文本两个方向哈希码长度k和哈希表个数m两个参数对模型效果的影响。实验是在NUS-WIDE和Wiki两个数据集上进行的,由于篇幅原因图2只给出了在Wiki上的实验结果。基于评价标准的定义,召回率越大越好,选择度越小越好。在NUS-WIDE数据集上的实验得到了类似的结果。

观察图2中任意一图,可以发现参数k和m对FCMR在两个检索方向上的影响趋势大致相



(a) Wiki数据集图像检索文本

(a) I2T on Wiki



(b) Wiki数据集文本检索图像

(b) T2I on Wiki

图2　哈希码长度与哈希表个数对模型的影响

Fig. 2　Evaluation of the different length of hash code for a varying number of hash tables

同:随着k增大,召回率和选择度逐渐减小;随着m的增大,召回率和选择度逐渐增大。对于不同的k,m的变化趋势有所变化,以图2(a)为例,k为1时,召回率和选择度基本呈比例增长,且增长步长变化较小;k为7时,随着m的增大,召回率逐渐比选择度增长加快,且二者增长幅度逐渐变小。通过对比模型在两个数据集上的表现,发现FCMR在Wiki上性能要更好,这是由于Wiki数据集一个图像对应长度平均为117.5个单词的文本,而NUS-WIDE数据集一个图像只对应长度平均为7.73个单词的文本,也就是说FCMR对于包含长文本的检索表现更好,这也在一定程度上说明了要得到满意的检索效果一般需要更多的细节信息。

将查询按其所属类别进行分类,并对比FCMR在每一类上的表现。权衡召回率和选择度,对于Wiki数据集选取$k=3$和$m=14$;对于NUS-WIDE数据集选取$k=4$和$m=20$。表2是FCMR在两个数据集两个检索方向上的各个评价指标的整体表现。实验结果证明FCMR模型得到的邻居列表中相关文件所占比例大约是原数据集中相关文件所占比例的2倍,图3和图4显示的实验结果也可以从将查询进行分类的角度说明这一点。表2还能

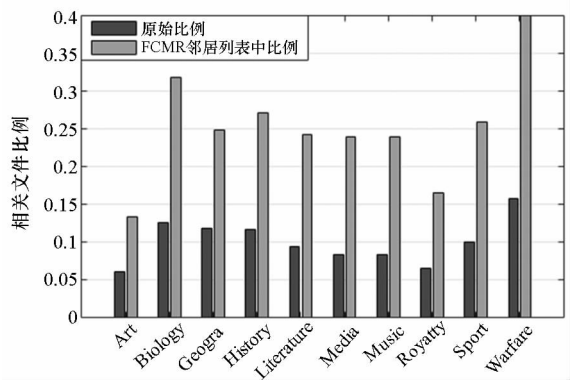
证明在两个数据集上,FCMR 在图像检索文本方向上的表现都优于文本检索图像方向的。同时在两个数据集上的实验结果证实了 FCMR 在 Wiki 数据集上的表现优于在 NUS-WIDE 数据集上的。

表 2　FCMR 实验结果

Tab. 2　Overall performance of FCMR

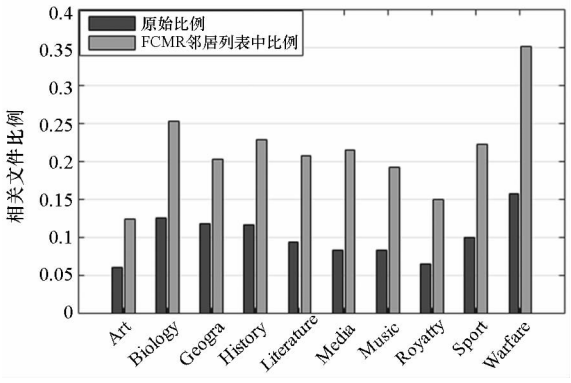
评价指标	Wiki		NUS-WIDE	
	I2T	T2I	I2T	T2I
召回率	0.862	0.850	0.733	0.733
选择度	0.326	0.386	0.404	0.425
相关文件比例提高	2.545	2.131	1.789	1.615

图 3 和图 4 从查询分类的角度分别显示了在 Wiki 和 NUS-WIDE 两个数据集上 FCMR 获得的邻居列表中相关文件所占比例与原测试数据集中相关文件所占比例的对比。通过图 3 可以发现 FCMR 在 Wiki 数据集上各个类别两个检索方向都有较为满意的表现。在图 4 中,只展示了具有代表性的 30 类数据在两个检索方向上的实验结果,从中可以发现 FCMR 在绝大多数类别中都表现较好,但是在“book”“garden”“glacier”和“computer”这四类中表现较差。对比这四类与其他类的原始数据的特征后发现,这四类中的大多数样本对应的文本描述只有很少的单词,主要原因是较少文本描述提供的检索信息不确定性大,降低了检索准确率。



(a) Wiki 数据集图像查询文本

(a) I2T on Wiki



(b) Wiki 数据集文本查询图像

(b) T2I on Wiki

图 3　FCMR 在 Wiki 数据集各类上相关文件比例表现

Fig. 3　Performance in terms of proportion over 10 concepts on Wiki

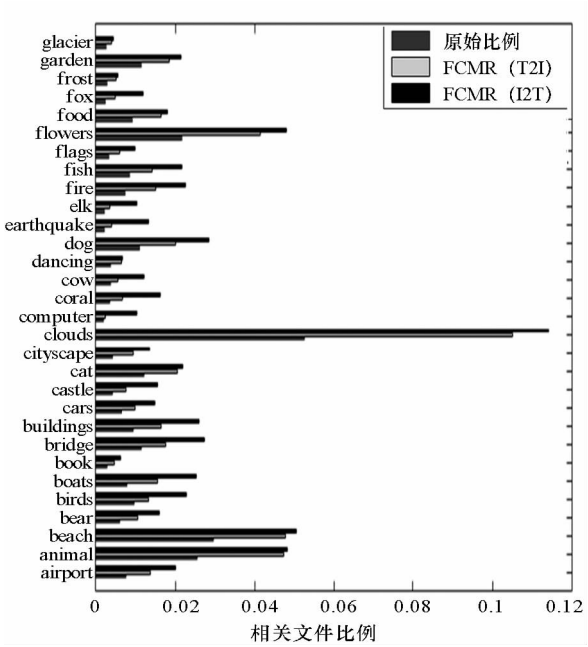


图 4　FCMR 在 NUS-WIDE 数据集各类上相关文件比例

Fig. 4　Performance in terms of proportion over 30 concepts on NUS-WIDE

最后,将 FCMR 与现有的两个跨媒体检索方法(CCA[4], Bi-CMSRM[8])结合,即使用 FCMR 模型获得邻居列表,然后用 CCA 和 Bi-CMSRM 在邻居列表中进行检索,并将检索结果与单独地使用 CCA,Bi-CMSRM 方法检索的进行比较。在实验中,对于 Wiki 数据集选取 $k=3$ 和 $m=14$;对于 NUS-WIDE 数据集选取 $k=4$ 和 $m=20$。最终用 MAP@ all 来评价检索效果,表 3 给出了 Wiki 数

表 3　各方法在 Wiki 和 NUS-WIDE 数据集上的 MAP@ all 表现

Tab. 3　Performance comparison in terms of MAP@ all scores on the Wiki and NUS-WIDE datasets

方法	Wiki		NUS-WIDE	
	T2I	I2T	T2I	I2T
CCA	0.143 3	0.145 1	0.085 1	0.088 3
FCMR-CCA	0.197 2	0.201 2	0.106 7	0.110 6
Bi-CMSRM	0.212 3	0.252 8	0.145 3	0.238 0
FCMR-Bi-CMSRM	0.284	0.305	0.186	0.296

据集和 NUS-WIDE 数据集的实验结果。从实验结果可以发现 FCMR 可以较大幅度提高原有方法的准确率,分析原因是应用 FCMR 模型显著地提高了邻居列表中与查询相关的文档所占比例。

5.3 潜在局限性及未来工作

虽然 FCMR 可以显著地提高邻居列表中与查询相关的文档所占比例从而提高检索效率及准确性,但是 LSH 从根本上来讲是一种概率方法,因此结果会有一定的波动。另外,前期对数据进行编码也需要大量的存储空间,从而限制了其处理大规模数据的能力。未来我们将在对 FCMR 生成的哈希码进行二次哈希并将方法应用于大数据集方面做一些有意义的工作。

6 结论

基于局部敏感哈希算法和神经网络学习,提出了一种新的跨媒体检索方法 FCMR,通过缩小检索范围,提高最终检索范围内相关文件所占比例,进而提高检索的准确率。在 Wiki 和 NUS-WIDE 两个数据集上检验了提出的方法,并将 FCMR 与现有的两个跨媒体检索方法 CCA 和 Bi-CMSRM 结合进行了进一步的实验,实验结果证实 FCMR 方法能够有效提高跨媒体检索的效率和准确率。

参考文献(References)

[1] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of Machine Learning research, 2003, 3: 993 – 1022.

[2] Blei D M, Jordan M I. Modeling annotated data [C]// Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003: 127 – 134.

[3] Jia Y, Salzmann M, Darrell T. Learning cross modality similarity for multinomial data [C]//Proceedings of IEEE International Conference on Computer Vision (ICCV), 2011: 2407 – 2414.

[4] Bian J, Yang Y, Zhang H W, et al. Multimedia summarization for social events in microblog stream[J]. IEEE Transactions on Multimedia, 2015, 17(2): 216 – 228.

[5] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods [J]. Neural Computation, 2004, 16 (12): 2639 – 2664.

[6] Ranjan V, Rasiwasia N, Jawahar C V. Multi-label cross-modal retrieval [C]//Proceedings of IEEE International Conference on Computer Vision, 2015: 4094 – 4102.

[7] Gong Y C, Ke Q D, Isard M, et al. A multi-view embedding space for modeling internet images, tags, and their semantics[J]. International Journal of Computer Vision, 2014, 106(2): 210 – 233.

[8] Wu F, Lu X Y, Zhang Z F, et al. Cross-media semantic representation via bi-directional learning to rank [C]// Proceedings of the 21st ACM International Conference on Multimedia, 2013: 877 – 886.

[9] Yan F, Mikolajczyk K. Deep correlation for matching images and text [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3441 – 3450.

[10] Yu T Y, Bai L, Guo J L, et al. A deep two-stream network for bidirectional cross-media information retrieval [C]// Proceedings of Pacific Rim Conference on Multimedia, 2016: 328 – 337.

[11] Kulis B, Grauman K. Kernelized locality sensitive hashing for scalable image search [C]//Proceedings of IEEE 12th International Conference on Computer Vision, 2009: 2130 – 2137.

[12] Li J Y, Li J H. Fast image search with locality-sensitive hashing and homogeneous kernels map [J]. The Scientific World Journal, 2015, 2015: 350676.

[13] Weems M A. Kernelized locality-sensitive hashing for fast image landmark association[M]. USA: Biblioscholar, 2011.

[14] Zhuang Y T, Yu Z, Wang W, et al. Cross-media hashing with neural net-works [C]//Proceedings of the ACM International Conference on Multimedia, 2014: 901 – 904.

[15] Kumar S, Udupa R. Learning hash functions for cross-view similarity search [C]// Proceedings International Joint Conference on Artificial Intelligence, 2011, 22 (1): 1360 – 1365.