中国科学:信息科学 2017年 第47卷 第8期:1109-1126

SCIENTIA SINICA Informationis

从大数据到大知识工程专刊・论文



# 基于多特征多核哈希学习的大规模图像检索

曾宪华<sup>1,2\*</sup>,袁知洪<sup>1,2</sup>,王国胤<sup>1,2</sup>,杨洁<sup>1,2</sup>

1. 重庆邮电大学计算机科学与技术学院, 重庆 400065

2. 重庆邮电大学计算智能重庆市重点实验室, 重庆 400065

\* 通信作者. E-mail: zengxh@cqupt.edu.cn

收稿日期: 2017-06-24; 接受日期: 2017-08-01; 网络出版日期: 2017-08-23 国家重点研发计划云计算和大数据重点专项 (批准号: 2016YFB1000905)、国家自然科学基金项目 (批准号: 61672120) 和重庆自然 科学基金项目 (批准号: cstc2015jcyjA40036) 资助

**摘要** 哈希学习方法解决了图像大数据的检索效率低,存储代价高的问题.目前已有的核哈希方法 中,要么仅使用一种特征对应单个核函数,要么是多特征对应单个核函数,它们忽视了综合考虑不同 的核函数具有的不同作用和不同的特征包含不同的信息的事实.本文提出了一种自适应的多特征多 核的哈希学习算法 (MFMKH),该算法能够自适应学习多特征融合的权重系数和多核融合的权重系 数,将多特征和多核的优点进行了双重融合.本算法中的特征融合解决了单特征所包含的信息量单 一不足的问题,采用多种不同的核函数能够弥补单核学习能力上的不足,具有多特征自适应融合和 多核学习的双重优点.在标准的 IRMA, Ultrasound 和 Cifar10 数据集上的实验表明,本文算法检索 性能明显优于同类基于核的哈希学习方法,且与监督的深度哈希相比训练时间显著少的情况下检索 性能在 Cifar10 数据集上是可竞争的.

关键词 维度约减,多特征融合,多核学习,哈希学习,自适应学习,图像检索

# 1 引言

在大数据时代,大数据的应用所涉及的领域越来越广.由于大数据的重要性,它已经成为国家重要的战略资源,对大数据的存储、管理和分析也已经成为学术界和工业界高度关注的热点<sup>[1,2]</sup>.而图像作为大数据信息来源的主要成分,它已成为视觉领域研究的热点.而机器学习技术<sup>[3]</sup>对于解决大规模图像检索中数据量大、维度高等问题起着关键的作用.

传统的图像检索分为基于文本的图像检索 (TBIR)<sup>[4]</sup> 和基于内容的图像检索 (CBIR)<sup>[5]</sup>.如今 CBIR 是当前图像检索的研究重点,对于 CBIR 而言,建立低层次图像特征和高层次的语义特征之间 的联系是很困难的,检索精度也相对较低.为此, Shao 等<sup>[4]</sup> 在 2004 年提出了一种基于文本和内容的 医学图像检索方法,该方法能够将高层次的语义特征和低层次的图像特征结合起来,使得检索效果比

引用格式: 曾宪华, 袁知洪, 王国胤, 等. 基于多特征多核哈希学习的大规模图像检索. 中国科学: 信息科学, 2017, 47: 1109–1126, doi: 10.1360/N112016-00307
 Zeng X H, Yuan Z H, Wang G Y, et al. Large-scale image retrieval based on multi-feature and multi-kernel hashing learning (in Chinese). Sci Sin Inform, 2017, 47: 1109–1126, doi: 10.1360/N112016-00307

© 2017《中国科学》杂志社

仅仅基于内容的图像检索方法要好. He 等<sup>[6]</sup> 在 2006 年提出一种基于广义的流形排序的图像检索方 法, 通过人工反馈的方法进一步提高了图像检索的精度. Wang 等<sup>[7]</sup> 在 2008 年提出一种基于内容的非 对等的流形排序图像检索方法,通过增加可靠数据点的权重,避免噪声的影响从而使得算法更具鲁棒 性. 然而上述提到的图像检索方法中都不能很好地应用于大规模的图像检索中, 当图像数据规模足够 大时,它们都面临着存储空间大,检索速度慢等缺点,因此哈希学习方法应运而生,其中一个最著名的 哈希方法就是局部感知哈希 (LSH)<sup>[8]</sup>,该方法由于采用的是随机映射的方式,因此不能很好地解决线 性不可分问题以及"维度灾难"问题.为了解决这个问题,Kulis等<sup>[9]</sup>在 2009 提出了基于核的局部敏 感哈希 (KLSH), 将其扩展为能够处理高维空间中的数据; 紧接着 Raginsky 等<sup>[10]</sup> 提出了平移不变性 核局部敏感哈希 (SKLSH); 以及 Liu 等 [11] 在 2012 年提出一种有监督的核哈希方法 (KSH), 该方法能 够巧妙处理哈希码的内积,从而可以更有效地优化汉明距离;考虑到多特征融合的优势,Liu 等<sup>[12]</sup>在 2014 年提出一种多特征核哈希方法 (MFKH). Li 等<sup>[13]</sup> 在 2015 年提出一种基于标签对的深度监督哈 希 (DPSH), 使用标签对信息使得该深度哈希方法能够同时进行特征学习和哈希编码学习. Liu 等<sup>[14]</sup> 在 2016 年提出一种深度监督哈希方法 (DSH),该方法使用卷积神经网络模型,使得最终学到紧凑的相 似性保留的哈希码. Shi 等<sup>[15]</sup> 在 2016 年提出一种基于核的离散监督哈希 (KSDH), 使用非对称松弛 策略,使得哈希函数学习和线性函数松弛能够同步进行,从而减少累计的量化误差.在大规模图像检 索中,大多数方法的计算时间主要是映射过程,为了能够尽可能地减少存储空间和降低计算代价,Xu 等<sup>[16]</sup> 在 2017 年提出一种监督稀疏哈希 (SSH), 使用稀疏正则化方法减少映射矩阵的参数数量, 避免 过拟合问题. Gui 等<sup>[17]</sup> 在 2017 年提出一种快速监督离散哈希 (FSDH), 采用将训练样本的类别标签 回归到相应的哈希码的策略来加速算法,使得 FSDH 不仅速度快,而且具有很好的检索性能.

哈希学习作为近年来大数据学习的一个研究热点<sup>[18]</sup>,它能够将数据表示成二进制码的形式,不仅能显著减少数据的存储和通信开销,还能降低数据维度,从而显著提高大数据学习系统的效率<sup>[19]</sup>.近几年,哈希方法在许多计算机视觉<sup>[20,21]</sup>任务中受到广泛的重视的主要原因在于哈希学习能够降低数据维度,减少存储空间,加快检索速度.

尽管前面提到的这些基于核哈希的方法<sup>[9~12]</sup> 在视觉检索中取得了很大的进展,但是已有的基于 核的哈希方法由于在学习哈希函数的过程中只是利用了某个单一的核函数以及某种特征而使其学习 能力受到限制.而现有的多特征哈希方法<sup>[22,23]</sup> 通过融合多种特征虽然已经取得了很好的效果,但是 在处理这些多特征的过程中,多特征串联成一种特征一方面会导致特征维度非常高,增加计算量;另 一方面无法发现不同特征之间的内在联系.它们都没有综合考虑不同的特征包含着不同的图像信息以 及不同的核函数有不同的作用与适用领域<sup>1)</sup>.

对于上述存在的问题,本文提出一种自适应的多特征多核的哈希学习方法,该方法先将每一种特征映射到不同的核空间中,然后将这些高维核空间的特征进行融合构造出一种多特征多核的哈希学习方法.在该方法中,不同特征和不同核都有不同的权重,这种方法的思想结合了多视图局部线性嵌入方法<sup>[24]</sup>和多核学习方法<sup>[25,26]</sup>,该方法具有多特征和多核的双重优点.本文的主要创新点在于:

(1) 使用基于学习的方法来学习线性组合的核函数中的每一个核函数的比重系数, 而非人为设定, 从而使该方法中每个核函数的选择具有更高的可信度.

(2) 使用基于学习的方法来学习线性组合中的每一种特征的比重系数, 从而能够更好地学习到特征之间的关联性以及每种特征的重要程度.

(3) 采用的是自适应的学习方法,从而能够自动选择性能较优的核以及效果较好的特征.

<sup>1)</sup> https://github.com/zhihongyuan/MFMKH.git.

# 2 相关工作

目前,由于医院里每天会产生大量的医学图像,这使得医学图像诊断的相关工作人员的工作压力 非常大.虽然现有的 CBIR 系统在一定程度上缓解了工作人员的压力<sup>[24,27]</sup>,但是该系统无法应用于大 规模的医学图像的检索中,而且其精度也有待提高.随着基于哈希学习的图像检索技术的出现,该问 题逐渐得到解决.然而这些基于哈希的方法<sup>[9~11]</sup> 也受到了一些限制,使得哈希函数的学习仅仅是通 过单一特征或单一的核函数得到.而多特征的哈希学习方法<sup>[22,23]</sup> 通过融合不同的特征虽然取得了很 好的效果,但是这些方法是将这些特征预先串联在一起作为一种特征向量或者是通过简单的线性组合 构成一种特征向量.一方面无法充分利用不同的特征之间的作用,另一方面这种将多特征连在一起的 方法使得特征维度增大,从而在训练和检索时,计算代价增大.因此将这些方法应用于大规模的维度很 高的视觉图像中是不可行的<sup>[12]</sup>.

谱哈希 SH<sup>[28]</sup> 方法的基本思想是将原始数据编码成紧凑的二进制哈希码的同时并在汉明空间中保证原始数据的相似性,该方法可以看作是谱聚类<sup>[29]</sup> 方法的一种扩展.它的目标函数如下:

$$\min \sum_{i,j} S_{ij} ||y_i - y_j||^2 \quad \text{s.t.} \quad y_i \in \{-1, 1\}^b, \quad \sum_i y_i = 0, \quad \frac{1}{n} \sum_i y_i y_i^{\mathrm{T}} = I, \tag{1}$$

其中 *S<sub>ij</sub>* 表示样本 *i* 和 *j* 的相似度, *y<sub>i</sub>* 表示第 *i* 个样本的哈希码, *b* 代表哈希码的 Bit 数. 由于式 (1) 中问题的求解是一个 NP 难题, 通过谱松弛后将离散限制条件转化为连续的就可以通过图的 Laplace 矩阵特征值分解方法求得最终的解.

另一个相关工作就是多特征核哈希<sup>[12]</sup>,它的基本思想是不同特征对应同一种核,然后再将这些映 射后的特征线性组合起来构成一种融合的特征.它的目标函数如下:

$$L(S, W, b, \mu) = \sum_{i,j=1}^{N} S_{ij} ||Y_i - Y_j||^2 + \lambda ||V||_F^2,$$
  
s.t.  $Y_i \in \{-1, 1\}^P, \quad \sum_{i=1}^{N} Y_i = 0, \quad (1/N) \sum_{i=1}^{N} Y_i Y_i^{\mathrm{T}} = I, \quad 1^{\mathrm{T}} \mu = 1, \quad \mu >= 0,$  (2)

其中 S<sub>ij</sub> 表示样本 i 和 j 的相似度, N 和 P 分别表示训练样本数与每个样本的哈希码数, Y<sub>i</sub> 表示二 进制哈希码矩阵 Y 中的第 i 列即第 i 个样本的哈希码.

从上述论述中可以发现,多特征核哈希<sup>[12]</sup>只利用核方法将不同特征融合在一起,并没有考虑到 同一种特征被映射到不同的核空间后再融合的效果.考虑到不同核函数的不同作用,不同特征包含不 同的信息,本文将不同特征,不同核函数进行线性组合,构造一种双重的多特征多核方法,使得它能够 融合任意特征以及任意核函数之间的双重组合.接下来将主要介绍本文多特征多核哈希学习方法.

## 3 自适应多特征多核哈希学习

在己有的基于核的哈希学习方法中,要么仅使用一特征对应一核如 KLSH<sup>[9]</sup>, SKLSH<sup>[10]</sup> 和 KSH<sup>[11]</sup>,要么就是多特征中的每一特征对应同一个核函数如 MFKH<sup>[12]</sup>.这些方法中并没有充分考虑 到不同特征所含有的不同信息以及不同核函数所具有的不同作用.

结合多特征核哈希<sup>[12]</sup>以及多核学习<sup>[25,26]</sup>的两种思想,本文提出一种自适应的多特征多核哈希 学习方法,其哈希学习示意图如图 1 所示.图 1 主要阐述了多特征多核哈希学习方法的流程:首先,对



图 1 (网络版彩图) 多特征多核哈希学习示意图

 ${\bf Figure \ 1} \quad {\rm (Color \ online) \ The \ diagram \ of \ multi-feature \ and \ multi-kernel \ hashing \ learning}$ 

于图像库中的每一幅图像,提取 M<sub>1</sub> 种不同的特征,然后将提取的每一种特征映射到 M<sub>2</sub> 个不同的核 空间中,接着将映射到核空间的特征进行组合构造出多特征多核空间,然后进行哈希学习,最后将学 习到的哈希码保存起来.本文提出一种自适应的多特征多核哈希学习方法,其多特征多核双重组合结 构图如图 2 所示,图 2 表明将每一种特征分别映射到不同的核空间中,然后将这些映射到核空间中的 特征进行组合,构造一种双重组合的多特征多核空间,最后在组合后的核空间中学习哈希函数.本文 提出的自适应的多特征多核哈希学习方法不仅能够解决单个特征包含的信息量不足问题,而且能够将 不同的核函数具有不同功能考虑进来,又因为核函数的特性,并不会因为特征的维度增加而导致"维 度灾难"问题.其中特征的权重系数和核函数的权重系数是能够自学习的而不是人为指定的,因此具 有更高的可信度.



图 2 多特征多核哈希学习结构图 Figure 2 The structure of multi-feature and multi-kernel hashing learning

# 3.1 目标函数的构造

本文令训练集的样本数为 N, 特征图的个数为  $M_1$ , 核函数的个数为  $M_2$ . 训练集的第 n 个样本的 第 m 种特征可以表示为  $X_n^{(m)} \in \mathbb{R}^{d_m \times 1}$ , 则  $X^{(m)} = [X_1^{(m)}, X_2^{(m)}, \dots, X_N^{(m)}] \in \mathbb{R}^{d_m \times N}$  表示所有训练样 本的第 m 种特征. 在第二部分相关工作的基础上, 构造多特征多核表达形式如下:

$$K(x,y) = \sum_{m=1}^{M_1} \mu_m \sum_{t=1}^{M_2} \alpha_t K_t(x,y),$$
(3)

其中  $\alpha_t$  表示第 t 个核函数的权重,  $\mu_m$  表示第 m 种特征的权重系数,  $K_t(x, y)$  表示图像 x 和 y 的第 t 个核函数的函数值. 由式 (3) 可知  $K = \sum_{m=1}^{M_1} \mu_m \sum_{t=1}^{M_2} \alpha_t K_t$ . 定义核映射函数如下:

$$\varphi(X_i) = \left[\mu_1^{1/2} \varphi_1^{\mathrm{T}} \left(X_i^{(1)}\right), \dots, \mu_j^{1/2} \varphi_j^{\mathrm{T}} \left(X_i^{(j)}\right), \dots, \mu_{M_1}^{1/2} \varphi_{M_1}^{\mathrm{T}} \left(X_i^{(M_1)}\right)\right]^{\mathrm{T}}, \tag{4}$$

该式中  $\varphi_j(X_i^{(j)})$  是第 *i* 个训练样本的第 *j* 种特征通过映射函数  $\varphi(\cdot)$  后的高维表达形式,  $\mu_j^{1/2}$  表示第 *j* 种特征所占的权重. 第 *p* 个哈希函数的线性映射定义如下:

$$h_p(\cdot) = \operatorname{sign}\left(V_p^{\mathrm{T}}\varphi(\cdot) + b_p\right), \quad t = 1, \dots, P,$$
(5)

其中第 *i* 个样本的第 *p* 个哈希码为  $Y_{pi} = h_p(X_i)$ .

定义超平面向量 Vq 可由 R 个映射到对应的核空间中的 Zr 基点表示而成,即

$$V_q = \sum_{r=1}^R W_{rq} \varphi(Z_r), \quad r = 1, \dots, R,$$
(6)

其中 W 是一个 R×B 的权重矩阵, B 为哈希码长度, 基点或锚点可以通过随机样本<sup>[30]</sup> 或者聚类<sup>[31]</sup> 的方法获得.由上述式子可知所有训练样本的哈希码 Y 可以写成如下核形式, 即

$$Y_j = \operatorname{sign}\left(W^{\mathrm{T}}K_j + b\right),\tag{7}$$

其中  $K_j$  为 R 个基点与 N 个样本点构成的大小为  $R \times N$  的核矩阵  $K_{R \times N}$  的第 j 列, 偏置  $b = [b_1, b_2, \dots, b_B]$ , B 为哈希码长度.

对于所有训练样本, 令需要学习的哈希函数的个数为 *B*, 哈希函数表示为 {*h*<sub>1</sub>,...,*h*<sub>*B*</sub>}, 学习到的 哈希码为 *Y* (*B* × *N* 的矩阵) 并且哈希码能够保持图像之间的某种相似性 *S* (稀疏矩阵), 令 *S*<sub>*ij*</sub> 表示 任意两个样本点 *i* 和 *j* 之间的相似度, 对于多特征多核中的监督方法而言, 相似度矩阵 *S* 定义如下:

$$S_{ij} = \begin{cases} 1, & i \text{ and } j \text{ is closed and similar,} \\ -1, & i \text{ and } j \text{ is dissimilar,} \\ 0, & \text{others.} \end{cases}$$
(8)

在谱哈希<sup>[28]</sup> 和多特征核哈希<sup>[12]</sup> 的基础上,基于上述所列的式子,本文构造出如下的一种自适应 的多特征多核哈希学习的目标函数:

$$\min_{W,b,\mu,\alpha} L(S,W,b,\mu,\alpha) = \sum_{i,j=1}^{N} S_{ij} ||Y_i - Y_j||^2 + \lambda \operatorname{Tr}(W^{\mathrm{T}}KW),$$
s.t.  $Y_i = \operatorname{sign}(W^{\mathrm{T}}K_i + b), \quad Y_i \in \{-1,1\}^B, \quad \sum_{i=1}^{N} Y_i = 0, \quad \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^{\mathrm{T}} = I,$ 

$$K = \sum_{m=1}^{M_1} \mu_m \sum_{t=1}^{M_2} \alpha_t K_t, \quad 1^{\mathrm{T}}\mu = 1, \quad \mu \ge 0, \quad 1^{\mathrm{T}}\alpha = 1, \quad \alpha \ge 0.$$
(9)

使用 Laplace 矩阵 L = D - S (其中 D 为对角矩阵,  $D_{ii} = \sum_{j=1}^{N} S_{ij}$ ). 目标函数就可以转化为如下形式:

$$L(S, W, b, \mu, \alpha) = \operatorname{Tr}(Y^{\mathrm{T}}LY) + \lambda \operatorname{Tr}(W^{\mathrm{T}}KW).$$
(10)

对于目标函数 (10), 通过优化权重矩阵 W, 权重系数  $\mu$  以及权重系数  $\alpha$ , 使得学习到的哈希函数 能够保留原始图像的相似性. 虽然将不同的特征映射到不同的核空间后再将它们连接起来会导致特征 维度非常高, 但是使用核技巧后最终的特征维度只有 R 维, 而且  $R \ll \sum_{m=1}^{M} d_m$ . 同现有的将多特征 串联成一个特征 <sup>[22, 23]</sup> 的方法相比, 计算量并不会因为特征维度的增加而显著增加.

#### 3.2 目标函数的求解方法

由于目标函数具有离散性和非凸性约束条件,使得它的求解过程非常困难.使用类似谱哈希<sup>[28]</sup>的 思想,将离散约束条件  $Y_i = sign(W^T K_i + b) \in \{-1,1\}^B$  松弛后就转化为  $Y_i = W^T K_i + b$ .通过谱松弛 后,目标问题就有了凸性质,也就是有了极值.因此,在优化求解过程中只需分几个步骤就可以求出最 优解.首先,固定系数  $\mu$  和  $\alpha$ ,求解 W, b 的过程可以通过特征值分解的方法. 然后固定 W, b,  $\alpha$  求解  $\mu$ ,最后求解  $\alpha$  时固定 W, b,  $\mu$ .

#### **3.2.1** 求解权重矩阵和偏置 W, b

固定  $\mu$  和  $\alpha$ , 由限制条件  $\sum_{i}^{N} Y_{i} = \sum_{i}^{N} (W^{T}K_{i} + b) = 0$  可得  $b = -\frac{1}{N}W^{T}K_{R\times N}1$ , 将 b 代入到  $Y_{i}$  中, 根据式 (6) 将 V 代入到目标函数 (10) 中, 目标函数可以重写为

$$L(S, W, b, \mu, \alpha) = \operatorname{Tr}\left(W^{\mathrm{T}}K_{R \times N} L K_{R \times N}^{\mathrm{T}}W\right) + \lambda \operatorname{Tr}(W^{\mathrm{T}}K_{R \times R}W) = \operatorname{Tr}(W^{\mathrm{T}}CW),$$

因  $(I - (1/N)11^{\mathrm{T}})(I - (1/N)11^{\mathrm{T}})^{\mathrm{T}} = I - (1/N)11^{\mathrm{T}}$ ,则约束条件  $(1/N)\sum_{i=1}^{N} Y_i T_i^{\mathrm{T}} = I$ 可以 变形为

$$\frac{1}{N}\sum_{i=1}^{N}Y_{i}Y_{i}^{\mathrm{T}} = \frac{1}{N}(W^{\mathrm{T}}K_{i}+b)(W^{\mathrm{T}}K_{i}+b)^{\mathrm{T}} = \frac{1}{N}W^{\mathrm{T}}K_{R\times N}\left(I-\frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)K_{R\times N}^{\mathrm{T}}W_{N}$$

令  $G = (1/N)K_{R \times N}(I - (1/N)11^{T})K_{R \times N}^{T}$ , 于是限制条件就转化为了  $W^{T}GW = I$ . 因此, W, b 的 求解就可以转化为

$$\min_{W} \operatorname{Tr}(W^{\mathrm{T}}CW) \quad \text{s.t.} \quad W^{\mathrm{T}}GW = I,$$
(11)

式 (11) 的求解过程可通过特征值分解方法 [29] 求得.

# **3.2.2** 求解特征权重系数 $\mu$

当固定 W, b 和 α, 目标函数 (10) 中特征权重系数 μ 求解过程如下:

$$\begin{split} L(S, W, b, \mu, \alpha) &= \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right) L \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right)^{\mathrm{T}} W \right) \\ &+ \lambda \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right) W \right) \\ &= \operatorname{Tr} \left( \sum_{m=1}^{M_{1}} \mu_{m} \mu_{m}^{\mathrm{T}} W^{\mathrm{T}} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} L \left( \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right)^{\mathrm{T}} W \right) + \lambda \operatorname{Tr} \left( \sum_{m=1}^{M_{1}} \mu_{m} W^{\mathrm{T}} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} W \right) \\ &= \mu^{\mathrm{T}} E \mu + h^{\mathrm{T}} \mu, \end{split}$$

其中  $E = \text{Tr}(W^{\mathrm{T}}(\sum_{t=1}^{M_2} \alpha_t K_t) L(\sum_{t=1}^{M_2} \alpha_t K_t)^{\mathrm{T}} W), h^{\mathrm{T}} = \lambda \text{Tr}(W^{\mathrm{T}}(\sum_{t=1}^{M_2} \alpha_t K_t) W).$ 于是目标函数 (10) 求 解系数可以转化为如下:

$$\min_{u} u^{\mathrm{T}} E u + h^{\mathrm{T}} u \quad \text{s.t.} \quad 1^{\mathrm{T}} u = 1, \quad u \ge 0,$$
(12)

因此求 μ 的过程可以看作是二次规划问题的求解过程.

## **3.2.3** 求解核函数权重系数 α

当固定 W, b 和 μ, 目标函数 (10) 中核权重系数 α 求解过程如下:

$$\begin{split} L(S, W, b, \mu, \alpha) \\ &= \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right) L \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right)^{\mathrm{T}} W \right) + \lambda \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{m=1}^{M_{1}} \mu_{m} \sum_{t=1}^{M_{2}} \alpha_{t} K_{t} \right) W \right) \\ &= \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{t=1}^{M_{2}} \alpha_{t} \sum_{m=1}^{M_{1}} \mu_{m} K_{t} \right) L \left( \sum_{t=1}^{M_{2}} \alpha_{t} \sum_{m=1}^{M_{1}} \mu_{m} K_{t} \right)^{\mathrm{T}} W \right) + \lambda \operatorname{Tr} \left( W^{\mathrm{T}} \left( \sum_{t=1}^{M_{2}} \alpha_{t} \sum_{m=1}^{M_{1}} \mu_{m} K_{t} \right) W \right) \\ &= \operatorname{Tr} \left( (\alpha_{1}, \alpha_{2}, \dots, \alpha_{M_{2}})^{\mathrm{T}} \sum_{m=1}^{M_{1}} \mu_{m} \left[ \begin{array}{c} W^{\mathrm{T}} K_{1} \\ W^{\mathrm{T}} K_{2} \\ \vdots \\ W^{\mathrm{T}} K_{M_{2}} \end{array} \right] \right) L \left( (\alpha_{1}, \alpha_{2}, \dots, \alpha_{M_{2}}) \sum_{m=1}^{M_{1}} \mu_{m} \left[ \begin{array}{c} K_{1}^{\mathrm{T}} W \\ K_{2}^{\mathrm{T}} W \\ \vdots \\ K_{M_{2}}^{\mathrm{T}} W \end{array} \right] \right) \end{split}$$

$$+\lambda \operatorname{Tr}\left((\alpha_{1}, \alpha_{2}, \dots, \alpha_{M_{2}})^{\mathrm{T}} \sum_{m=1}^{M_{1}} \mu_{m} \begin{bmatrix} W^{\mathrm{T}} K_{1} W \\ W^{\mathrm{T}} K_{2} W \\ \vdots \\ W^{\mathrm{T}} K_{M_{2}} W \end{bmatrix}\right) = \alpha^{\mathrm{T}} U \alpha + \alpha^{\mathrm{T}} Q,$$

其中  $\alpha^{\mathrm{T}} = (\alpha_1, \alpha_2, \dots, \alpha_{M_2})^{\mathrm{T}},$ 

$$U = \operatorname{Tr}\left(\left(\sum_{m=1}^{M_{1}} \mu_{m} \begin{bmatrix} W^{\mathrm{T}}K_{1} \\ W^{\mathrm{T}}K_{2} \\ \vdots \\ W^{\mathrm{T}}K_{M_{2}} \end{bmatrix}\right) L \left(\sum_{m=1}^{M_{1}} \mu_{m} \begin{bmatrix} K_{1}^{\mathrm{T}}W \\ K_{2}^{\mathrm{T}}W \\ \vdots \\ K_{M_{2}}^{\mathrm{T}}W \end{bmatrix}\right)\right), \quad Q = \lambda \operatorname{Tr}\left(\sum_{m=1}^{M_{1}} \mu_{m} \begin{bmatrix} W^{\mathrm{T}}K_{1}W \\ W^{\mathrm{T}}K_{2}W \\ \vdots \\ W^{\mathrm{T}}K_{M_{2}}W \end{bmatrix}\right).$$

于是通过目标函数 (10) 求解系数 α 可以简化如下:

$$\min_{\alpha} \alpha^{\mathrm{T}} U \alpha + \alpha^{\mathrm{T}} Q \quad \text{s.t.} \quad 1^{\mathrm{T}} \alpha = 1, \quad \alpha \ge 0,$$
(13)

式 (13) 的求解过程同样可通过二次规划方法求出 α.

#### 3.2.4 哈希编码过程

对于新的待检索样本 x, 它的哈希码表达形式为

$$y = \operatorname{sign}(W^{\mathrm{T}}[K(x, Z_1), \dots, K(x, Z_R)]^{\mathrm{T}} + b),$$
(14)

其中 Z<sub>i</sub> 表示 R 个基点中的第 i 个基点样本.使用汉明距离排序或者哈希查找表的方法,求出待检索的样本与数据库中的样本的距离,可以在次线性时间内检索出相似图像.

综上分析,本文的 MFMKH 算法步骤描述如算法 1:

算法 1 多特征多核哈希检索算法
训练过程:
<b>输入:</b> 训练集 $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\};$
输出: $W, b, \mu, \alpha;$
1. $\mu_j = \frac{1}{M_1}, \ j = 1, \dots, M_1 \ \Re \ \alpha_i = \frac{1}{M_2}, \ i = 1, \dots, M_2;$
repeat
<ol> <li>固定 α 和 μ, 求解 W, b 通过式 (11);</li> </ol>
<ol> <li>固定 W, b 和 α, 求解 μ 通过式 (12);</li> </ol>
4. 固定 W, b 和 μ, 求解 α 通过式 (13);
until 收敛.
测试过程:
<b>输入:</b> 测试集 $X_{\text{test}}$ 以及根据训练过程得到的 $W$ , $b$ , $\alpha$ , $\mu$ ;
输出:前 n 幅与待检索样本相似的图像;
1. 根据式 (14) 算出测试集的哈希码 Y <sub>test</sub> ;
2. 计算测试集哈希码 Y <sub>test</sub> 与数据库中所有的图像的哈希码的汉明距离向量 P;
3. 对汉明距离向量 P 计算出的距离进行升序排序,返回前 n 张图像.

# 4 算法分析

本节主要对提出的多特征多核哈希学习算法 (MFMKH) 进行算法复杂度分析和空间复杂度分析. 本算法通过随机样本法选取基点后能够降低时间复杂度,不会因为特征融合后的维度增加而导致计算 量增大,其复杂度只与选择的基点数和样本的规模有关.

#### 4.1 时间复杂度分析

本文提出的多特征多核哈希学习算法的时间复杂度由多特征多核哈希编码训练过程的时间复杂 度决定,训练时每次迭代的时间复杂度由 3 部分构成,分别是固定  $\mu$  和  $\alpha$  时,求解 W, b 的时间复 杂度;固定 W, b,  $\alpha$  时,求解  $\mu$  的时间复杂度以及固定 W, b,  $\mu$  时,求解  $\alpha$  的时间复杂度共同决 定.其中求解权重矩阵 W 和偏置 b 的时间复杂度主要是 3 个矩阵的乘积即计算  $K_{R\times N}LK_{R\times N}^{T}$  (其 中 L 是通过 k 近邻计算得到的每列只有 k 个非零元素的对称稀疏矩阵 —— Laplace 矩阵, k 为常数,  $K_{R\times N} = \sum_{m=1}^{M_1} \mu_m \sum_{t=1}^{M_2} \alpha_t K_t$ ),所以计算  $K_{R\times N}LK_{R\times N}^{T}$  需要进行  $k \times M_1 \times M_2 \times R \times R \times N$  次乘法 运算 (k 为常数),故其时间复杂度为  $O(M_1M_2R^2N)$ ;求解特征权重系数  $\mu$  的时间复杂度主要是计算  $(\sum_{t=1}^{M_2} \alpha_t K_t)L(\sum_{t=1}^{M_2} \alpha_t K_t)^{T}$ ,同求解 W, b 的时间复杂度类似,其时间复杂度为  $O(M_2^2R^2N)$ ;同理,求 解核函数权重系数  $\alpha$  的时间复杂度为  $O(M_1^2R^2N)$ .所以当迭代次数为 T 时,本算法的总时间复杂度 为  $O(TR^2N(M_1M_2 + M_1^2 + M_2^2))$ .

# 4.2 空间复杂度分析

哈希学习将数据映射成二进制串的形式, 能显著减少数据的存储和通信开销, 从而有效提高学习 系统的效率. 训练过程, 不考虑输入的训练集存储空间, 只考虑新引入的参数存储空间, 则需保存的参 数有权重矩阵 W, 偏置 b, 特征权重系数  $\mu$ , 核权重系数  $\alpha$  以及输出训练样本的哈希码 Y, 整个训练过 程中的空间复杂度由这 5 个部分组成. W 的空间复杂度为 O(RB); b 的空间复杂度为 O(NB);  $\mu$  的 空间复杂度为  $O(M_1)$ ,  $\alpha$  的空间复杂度为  $O(M_2)$ , Y 的空间复杂度为 O(BN). 其中 R 为选择的基点 数, B 为训练样本的哈希码长度, N 为训练样本数,  $M_1$  为特征个数,  $M_2$  为核函数个数. 因此整个训练 过程的空间复杂度为  $O(RB + NB + M_1 + M_2)$ . 存储过程, 每个数据点会被一个紧凑的二进制串编码, 在原空间中相似的 2 个点应当被映射到哈希码空间中相似的 2 个点. 对于总样本数为 N, 总特征维度 为 M, 每一个维度所需的存储空间为  $\delta$  byte, 所需总的存储空间为  $\delta \times N \times M$  byte, 而如果把每个数 据映射成哈希码, 则所需的存储空间为  $(N \times B)/8$  byte, 其中 B 为每个样本的哈希码长度  $(B \ll M)$ , 因此存储空间复杂度为 O(NB).

#### 4.3 算法收敛性分析

在本文的 3 个实验中, 通过交替迭代的方式优化目标函数 (10), 对于每一次交替迭代进行 3 步更 新, 第 1 步先固定  $\alpha$  和  $\mu$  更新 W 和 b, 第 2 步固定 W, b 和  $\mu$  更新  $\alpha$ , 第 3 步固定 W, b 和  $\alpha$  更新  $\mu$ , 交替更新直到目标函数 (10) 的函数值稳定时. 实验过程中, 设目标函数 (10) 前后两次迭代的误差 为 error, 当 error 小于 1E-4 时, 算法收敛. 在 IRMA, Ultrasound, Cifar10 数据集上的实验结果表明 当哈希码长度分别为 8, 12, 32, 24, 48 bits 时, 目标函数在 IRMA 数据集上分别迭代 82, 45, 37, 55, 42 次后算法收敛; 目标函数在 Ultrasound 数据集上分别迭代 28, 85, 89, 74, 38 次后算法收敛; 目标函数 在 Cifar10 数据集上分别迭代 291, 243, 142, 128, 102 次后算法收敛. 图 3(a) ~ (c) 分别表示目标函数



图 3 (网络版彩图)误差与迭代次数关系

Figure 3 (Color online) Error vs. iterations on (a) IRMA dataset, (b) Ultrasound dataset, and (c) Cifar10 dataset

收敛时在 3 个不同数据集上误差与迭代次数的关系 (不同数据集下不同方法的训练时间表请查阅网 上附件<sup>2)</sup>).

# 5 实验结果

本节将本文提出的方法与相关方法进行实验比对,从而验证本文提出方法的有效性. 主要比对的 方法有核局部感知哈希<sup>[9]</sup>、平移不变性核局部感知哈希<sup>[10]</sup>、多特征核哈希<sup>[12]</sup>、谱哈希<sup>[28]</sup>以及基于 标签对的深度监督哈希<sup>[13]</sup>.此外,本文还进行了自我对比的实验即多特征多核实验中将任意单特征 单核的组合形式和多特征多核方法进行对比. 实验过程中,为了减少随机性的影响,所有的方法进行 了 10 次实验.

本文实验中使用 GIST 特征、HOG 特征、线性核、Laplace 核及 Gauss 核的组合包括理论分析和 多次实验验证.

(1)本文从特征和核函数的物理特性进行分析说明选择该组合的合理性,由于特征的选择多种多 样,本文主要从全局特征和局部特征进行考虑.全局特征主要考虑的是图像的宏观特征而忽略图像的 局部特征 (如 GIST 特征);而局部特征是图像特征的局部表达,它只能反应图像上具有的局部特殊性 (如 HOG 特征).核函数主要分为全局核函数和局部核函数,全局核函数 (如线性核函数)具有全局特 性,其允许相距很远的数据点对核函数的值有影响,而局部核函数具有局部性,只允许相距很近的数 据点对核函数的值有影响.Laplace 核作为一种局部核只保证数据在原始空间中是近邻关系的点通过 映射后依然保持近邻关系,而没有保证非近邻的点通过映射后尽可能的保持远,这就导致非近邻的点 通过映射后可能变成了近邻点.Gauss 核表示两个样本点的距离,它能够保证非近邻点通过映射后还 是非近邻关系,弥补了 Laplace 核函数的不足.因此从特征的全局性和特征的局部性,以及核函数的全

<sup>2)</sup> https://github.com/zhihongyuan/MFMKH.git.

局特性和局部特性考虑,本文选择了这两种特征3种核函数的结合方法.

(2) 通过实验表明,将 HOG 特征和 GIST 特征与线性核、Gauss 核和 Laplace 核进行组合时效果 比仅仅将 HOG 特征或者 GIST 特征单独与 3 种核函数组合时效果要好.

(3) 未来将其他手工特征和深度特征结合探索更有利于实际应用场景的特征融合,并探索更多的 核函数,总体上应该没有适合一切应用问题的某一个或几个特征,也没有适合一切应用问题的一种或 几种核函数度量,都需要根据实验场景进行具体分析与测试.

本文主要用两种常用的检索方法来评估哈希学习的性能: 汉明排序和哈希查找表. 用哈希算法进 行检索主要就是利用了汉明空间离散的特点, 在汉明空间中可以用机器指令或查表来计算距离 (减法), 相比于欧氏距离计算其速度相当快, 有几十甚至上百倍的差距. 汉明排序是将待检索样本同数据库中 所有的样本点的汉明距离按升序排序, 而哈希查找表使用的是二进制哈希码查找方法, 将所有的查询 结果中汉明距离小于 2 的样本返回作为查询结果.

#### 5.1 数据集

本文在 IRMA 数据集<sup>3)</sup>、Ultrasound 数据集<sup>4)</sup> 以及 Cifar10<sup>5)</sup> 数据集上进行了实验验证,实验过 程中本文主要采用两种特征和 3 种核的组合方式.其中特征为 HOG 特征和 GIST 特征,核函数为线 性核、Laplace 核和 Gauss 核. IRMA 数据集包含了 193 个类共 12677 张有注释的灰度射线图.其中 有 57 个类别共 10902 张图具有标签信息,由于 IRMA 数据集中只有 2005 年的数据集是有标签的数 据,其他为无标签的数据,而本文中采用的是有监督的方法,所以选取了 IRMA 数据集中 2005 年的数 据. 该数据集共 57 类 10000 张,由于该数据集中每个类别的数量参差不齐,大部分类别中数量低于 50 张,而为了能够更好地训练模型,本文选取的每种数据样本集在 150 张以上,而满足要求的只有 15 种, 从这 15 种中随机选取 10 类共 2673 张,对于每个类别大约按 3:1 的比例随机分训练样本和测试样本, 具体分布情况如表 1.

Ultrasound 数据集总量为 2682 分 10 个类别,每个类大约按 3:1 的比例随机选取样本作为训练集 和测试集,具体分布情况如表 1. 对于 Cifar10 数据集,本文随机选取 1000 张 (每类 100 张) 作为测试 集,从剩余的图像中随机选择 5000 张 (每类 500 张) 作为训练集.

#### 5.2 实验分析

在 IRMA, Ultrasound 和 Cifar10 数据集上验证了本文提出的自适应多特征多核哈希方法 (MFMKH). 本文在 IRMA 数据集上验证了核哈希方法之间的准确率、召回率与返回检索样本数和 哈希码长度之间的关系. 另外, 本文又将核哈希和深度哈希在 MAP (mean average precision) 指标和训练时间做了对比实验.

核哈希对比方法中本文采用随机样本法从训练集中选择了 300 个样本作为基点,实验过程中采用 了 GIST<sup>[32]</sup> 特征和 HOG<sup>[33]</sup> 特征,其中 GIST 特征维度为 384 维,HOG 特征维度为 324 维;核函数采 用了 Gauss 核、Laplace 核和线性核.在核哈希和深度哈希的对比实验中,本文采用随机样本法从训练 集中选择了 300 个样本作为基点.实验过程中采用了 GIST 特征和 HOG 特征,其中 GIST 特征维度 为 512 维,HOG 特征维度为 324 维;核函数采用了 Gauss 核、Laplace 核和线性核.本文在上述 3 个数 据集中从 MAP 指标上对它们进行了比较分析.在这 3 个对比实验中,SH<sup>[28]</sup>,SKLSH<sup>[10]</sup>,KLSH<sup>[9]</sup> 对

<sup>3)</sup> http://www.irma-project.org.

<sup>4)</sup> Ultrasound 数据集为本团队收集的临床超声图像数据,来源于重庆市某医院.

<sup>5)</sup> http://www.cs.toronto.edu/ kriz/cifar.html.

IRMA				Ultrasound			
Class	Sum	Training	Testing	Class	Sum	Training	Testing
0	336	286	50	0 (Gallstone)	521	371	150
1	215	165	50	1 (Gallbladder polyps)	238	178	60
2	225	175	50	2 (normal Gallbladder )	154	104	50
3	576	476	100	3 (Hydronephrosis)	386	286	100
4	217	167	50	4 (Kidney stones)	219	169	50
5	205	155	50	5 (Renal cyst)	197	147	50
6	194	144	50	6 (Normal kidneys)	303	233	70
7	284	234	50	7 (Liver cysts)	113	83	30
8	228	178	50	8 (Fatty liver)	360	280	80
9	193	143	50	9 (Normal liver)	191	131	60

表 1 IRMA 和 Ultrasound 图像分布表 Table 1 The distribution of IRMA and Ultrasound images

应的特征为 512 维的 GIST 特征, MFKH<sup>[12]</sup>, M2FM3K 对应的特征为 512 维的 GIST 特征和 324 维的 HOG 特征的组合, DPSH<sup>[13]</sup> 输入的为原始图像. 在 IRMA, Ultrasound 和 Cifar10 数据集上, Gauss 核、Laplace 核和线性核的参数分别为 5, 1, 10; 0.5, 5, 1; 25, 0.1, 10.

本文主要对比的方法有多特征核哈希 (MFKH)<sup>[12]</sup>、谱哈希 (SH)<sup>[28]</sup>、平移不变核局部感知哈希 (SKLSH)<sup>[10]</sup>、核局部感知哈希 (KLSH)<sup>[9]</sup>以及基于标签对的深度监督哈希 (DPSH)<sup>[13]</sup>.

#### 5.2.1 核哈希方法的实验对比分析

本小节主要对核哈希方法在 IRMA 数据集上进行实验对比分析, 主要对比的方法有多特征核哈希 (MFKH)、谱哈希 (SH)、平移不变核局部感知哈希 (SKLSH)、核局部感知哈希 (KLSH).

图 4 和 5 是在 IRMA 数据集上进行的实验, 主要比较本文方法中的多特征多核方法 (M2FM3KH 中特征为 GIST 和 HOG) 与其他对比方法在返回检索样本数为 50 时, 不同长度的哈希码下准确率 与召回率的关系. 当返回检索样本数为 50, 哈希码长度大于等于 32 时, M2FM3KH 准确率超过了其 他对比方法, 而召回率仅次于谱哈希方法, 其中 M2FM3KH 中特征为 HOG 和 GIST, 核函数为线性 核、Laplace 核以及 Gauss 核, 核参数 σ 分别为 0.5, 100, 1 (此外, 本文还将单个 HOG 特征或 GIST 特征同上述 3 种不同的核函数进行了组合 (M1FM3KH) 并与其他核哈希方法进行了实验对比, 从实验 中可看出 M2FM3KH 实验结果比 M1FM3KH 实验结果好, 进一步验证了融合后的特征包含的图像信 息量更大, 能够进一步提高图像检索的准确率<sup>6</sup>).

图 6 和 7 是在 IRMA 数据集上进行的实验, 主要比较本文方法中的多特征多核方法 (M2FM3KH) 与该方法中的单特征单核组合方法 (共 6 种) 在返回样本数为 50 时准确率和召回率上的关系. 图 6 和 7 中的对比实验能够进一步验证本文提出的方法的可行性和优点. 从图 6 和 7 可以看出当哈希码长度 大于等于 128 时, M2FM3KH 的准确率和召回率超过了单特征单核的组合方法 (其中 M2FM3KH 中 特征为 HOG 和 GIST, 核函数为线性核、Laplace 核以及 Gauss 核, 核参数 σ 分别为 0.5, 20, 2; 其他 6 种对比方法为 2 种特征 3 种核函数的任意一种组合).

<sup>6)</sup> https://github.com/zhihongyuan/MFMKH.git.



™MFKH ■ SH SKLSH ■KLSH ■M2FM3KH



Figure 4 (Color online) The relationship based on kernel hashing methods between the length of the hash code and the accuracy rate when the number of retrieved samples is 50





Figure 5 (Color online) The relationship based on kernel hashing methods between the length of the hash code and the recall rate when the number of retrieved samples is 50



图 6 (网络版彩图) 在返回检索样本数为 50 时,基于 M2FM3KH 中的单特征单核的哈希码长度与准确率的关系 Figure 6 (Color online) The relationship between the length of the hash code and the accuracy rate based on single feature and single kernel hashing methods of M2FM3KH when the number of retrieved samples is 50



图 7 (网络版彩图) 在返回检索样本数为 50 时,基于 M2FM3KH 中的单特征单核的哈希码长度与召回率的关系 Figure 7 (Color online) The relationship between the length of the hash code and the recall rate based on single feature and single kernel hashing methods of M2FM3KH when the number of retrieved samples is 50

				-			× ×	·		
Method	IRMA (MAP)					Ultrasound (MAP)				
	8 bits	12  bits	24  bits	32  bits	48  bits	8 bits	12  bits	24  bits	32  bits	48 bits
MFKH	0.620	0.680	0.640	0.670	0.582	0.109	0.121	0.137	0.163	0.155
SH	0.214	0.315	0.323	0.416	0.409	0.063	0.061	0.054	0.052	0.056
SKLSH	0.128	0.126	0.121	0.122	0.147	0.174	0.203	0.231	0.283	0.206
KLSH	0.143	0.154	0.156	0.151	0.147	0.152	0.147	0.142	0.137	0.149
M2FM3KH	0.520	0.643	0.713	0.723	0.640	0.177	0.152	0.163	0.194	0.287
DPSH	0.896	0.969	0.971	0.968	0.971	0.524	0.527	0.582	0.590	0.601

表 2 IRMA 和 Ultrasound (MAP) 对照表 Table 2 Comparison of IRMA and Ultrasound (MAP)

#### 5.2.2 核哈希方法与深度哈希方法的实验对比分析

本小节主要对核哈希方法与深度哈希方法 (DSPH) 在 3 个不同数据集上的实验对比分析, 主要对 比的方法有多特征核哈希 (MFKH)、谱哈希 (SH)、平移不变核局部感知哈希 (SKLSH)、核局部感知 哈希 (KLSH) 以及基于标签对的深度监督哈希 (DPSH).

表 2 是在 IRMA 和 Ultrasound 数据集上进行的实验, 主要验证本文方法 (M2FM3KH) 与其他 对比方法包括深度哈希方法在不同哈希码长度下的 MAP. 从 IRMA 数据集上的实验结果中可以看 出虽然本文的方法整体的 MAP 要高于其他核方法, 但是同 DPSH 的 MAP 相比, 仍有所差距. 从 Ultrasound 数据集上的实验结果可以看出右边的 MAP 同左边的 MAP 相比要低很多, 进一步验证了 超声图像数据识别的难度. 表 3 是在 Cifar10 数据集上进行的实验, 从中可以看出 MFKH 和本文的 MFMKH 方法的 MAP 整体上比其他核方法高一点, 体现了多特征核方法和多特征多核方法比单特征 单核方法要好, 但依然低于 DPSH 的 MAP. 从这 3 个实验数据中可以看出本文方法虽然整体上比其 他核方法要好, 但是同 DPSH 相比, 仍有所差距, 体现了深度学习在图像检索方面的优势.

#### 5.3 实验总结

表 2 和 3 的实验结果表明本文算法检索性能虽然低于 DPSH 方法,但是明显优于同类基于核的 哈希学习方法.从表 2 和 3 可以看出 DPSH 的 MAP 整体上比核哈希方法要高出许多,但是 DPSH

1122

Table 5 Comparison of Charlo (MAF)								
Method			IRMA (MAP)					
	8 bits	12 bits	24 bits	32 bits	48 bits			
MFKH	0.235	0.261	0.252	0.241	0.247			
SH	0.104	0.116	0.145	0.142	0.139			
SKLSH	0.116	0.123	0.131	0.225	0.321			
KLSH	0.132	0.143	0.147	0.145	0.154			
M2FM3KH	0.231	0.264	0.367	0.423	0.243			
DPSH	0.681	0.713	0.727	0.744	0.757			

表 3 Cifar10 (MAP) 对照表 Cable 3 Comparison of Cifar10 (MAP)

的高 MAP 需要极大的时间代价. 虽然本文的方法在 MAP 上比 DPSH 方法低, 但本文方法充分利用 了多特征多核双重组合的优点, 通过多特征融合的方法来提高图像的信息量, 通过核方法来解决线性 不可分问题, 通过选择基点的方法使得其训练过程不受总特征维度的影响, 只与选择的基点个数和样 本规模有关. 与 DPSH 方法相比在训练时间显著减少的情况下检索性能在一定长度的哈希码上是可 竞争的.

### 6 结束语

本文提出了自适应多特征多核哈希学习 (MFMKH) 算法,该方法兼容基本的数据类型和各种相似 性度量方法,能够适用于多种形式的特征和多种核函数的双重组合方式,不存在数据需要具有某个分 布的提前假设,也不需要知道特征空间的映射形式.这种自适应的多特征多核的哈希学习算法架构能 够自适应学习特征权重系数和多核权重系数,将多特征和多核的优点进行了双重融合.多特征的融合 解决了单特征所包含的信息量单一不足的问题;多核的组合方式能够弥补单核学习能力上的不足且解 决了"维度灾难"问题.实验结果表明本文的方法具有多特征融合<sup>[24]</sup>和多核学习<sup>[25,26]</sup>的双重优点. 实验过程中,发现核函数的参数对实验结果有一定的影响,而且所提供的特征基本上是基于手工特征. 在后期的工作中,我们将 MFMKH 方法用于其他图像数据集上进行测试,进一步研究核参数的问题 以及深度特征与手工特征结合的效果.虽然本文方法在 MAP 上与 DPSH 相比有差距,但是本文方法 充分利用了多特征多核双重组合的优点,通过多特征融合的方法来提高图像的信息量,通过核方法来 解决线性不可分问题,通过选择基点的方法使得其训练过程不受总特征维度的影响,只与选择的基点 个数和样本规模有关.鉴于深度哈希学习的效果较好但所需的训练时间极长,在后期的工作中,我们 将会在本文的基础上进一步研究深度学习方法与本文的方法结合的可能性,使得训练时间能够进一步 减少.

#### 参考文献 —

Naimi A I, Westreich D J. Big data: a revolution that will transform how we live, work, and think. Am J Epidemiol, 2014, 17: 181–183

<sup>2</sup> Tu Z P. The Big Data Revolution. Guilin: Guangxi Normal University Press, 2013. 312–315 [涂子沛. 大数据. 桂林: 广西师范大学出版社, 2013. 312–315]

<sup>3</sup> Zhou Z H. Machine learning and data mining. Commun China Compute Fed, 2007, 3: 35-44 [周志华. 机器学习与数据挖掘. 中国计算机学会通讯, 2007, 3: 35-44]

- 4 Shao H, Cui W, Zhao H. Medical image retrieval based on visual contents and text information. In: Proceedings of International Conference on Systems, Man and Cybernetics, Hague, 2004. 1098–1103
- 5 Rui Y, Huang T S, Chang S. Image retrieval: past, present, and future. J Vis Commun Image Represent, 1999, 10: 1–23
- 6 He J, Li M, Zhang H J, et al. Generalized manifold-ranking-based image retrieval. IEEE Trans Image Process, 2006, 15: 3170–3177
- 7 Wang F, Er G, Dai Q. Inequivalent manifold ranking for content-based image retrieval. In: Proceedings of International Conference on Image Processing, California, 2008. 173–176
- 8 Har-Peled S, Indyk P, Motwani R. Approximate nearest neighbor: towards removing the curse of dimensionality. Symp Theory Comput, 2012, 52: 604–613
- 9 Kulis B, Grauman K. Kernelized locality-sensitive hashing for scalable image search. In: Proceedings of International Conference on Computer Vision, Kyoto, 2009. 2130–2137
- 10 Raginsky M, Lazebnik S. Locality-sensitive binary codes from shift-invariant kernels. In: Proceedings of Advances in Neural Information Processing Systems, Vancouver, 2009. 1509–1517
- 11 Liu W, Wang J, Ji R, et al. Supervised hashing with kernels. In: Proceedings of Conference on Computer Vision and Pattern Recognition, Providence, 2012. 2074–2081
- 12 Liu X, He J, Lang B. Multiple feature kernel hashing for large-scale visual search. Pattern Recog, 2014, 47: 748–757
- 13 Li W J, Wang S, Kang W C. Feature learning based deep supervised hashing with pairwise labels. Comput Sci, 2015,
   2: 1711–1717
- 14 Liu H, Wang R, Shan S, et al. Deep supervised hashing for fast image retrieval. In: Proceedings of Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 2064–2072
- 15 Shi X S, Xing F Y, Cai J Z, et al. Kernel-based supervised discrete hashing for image retrieval. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 419–433
- 16 Xu Y, Shen F, Xu X, et al. Large-scale image retrieval with supervised sparse hashing. Neurocomputing, 2017, 229: 45–53
- 17 Gui J, Liu T L, Sun Z N, et al. Fast supervised discrete hashing. IEEE Trans Pattern Anal Mach Intel, 2017. doi: 10.1109/TPAMI.2017.2678475
- 18 Li W J, Zhou Z H. Learning to hash for big data: current status and future trends. Sci Bull, 2015, 60: 485–490 [李武 军, 周志华. 大数据哈希学习: 现状与趋势. 科学通报, 2015, 60: 485–490]
- 19 Liu X L, Huang L, Deng C, et al. Multi-view complementary hash tables for nearest neighbor search. In: Proceedings of International Conference on Computer Vision, Santiago, 2015. 1107–1115
- 20 Xu H, Wang J, Li Z, et al. Complementary hashing for approximate nearest neighbor search. In: Proceedings of International Conference on Computer Vision, Barcelona, 2011. 1631–1638
- 21 Kan M, Xu D, Shan S, et al. Semisupervised hashing via kernel hyperplane learning for scalable image search. IEEE Trans Circ Syst Video Technol, 2014, 24: 704–713
- 22 Zhang D, Wang F, Si L. Composite hashing with multiple information sources. In Proceeding of International Conference on Research and Development in Information Retrieval, Beijing, 2011. 225–234
- 23 Song J, Yang Y, Huang Z, et al. Multiple feature hashing for real-time large scale near-duplicate video retrieval. In: Proceedings of International Conference on Multimedea, Scottsdale, 2011. 423–432
- 24 Shen H, Tao D, Ma D. Multiview locally linear embedding for effective medical image retrieval. Plos One, 2013, 8: e82409
- 25 Gonen M, Alpaydin E. Multiple kernel learning algorithms. J Mach Learn Res, 2011, 12: 2211–2268
- 26 Tzortzis G, Likas A. Greedy unsupervised multiple kernel learning. In: Proceedings of Hellenic Conference on Artificial Intelligence, Berlin, 2012. 73–80
- Akgül C B, Rubin D L, Napel S, et al. Content-based image retrieval in radiology: current status and future directions.
   J Digit Imag, 2011, 24: 208–222

- 28 Weiss Y, Torralba A, Fergus R. Spectral Hashing. In: Proceedings of Conference on Neural Information Processing Systems, Auckland, 2008. 1753–1760
- 29 Zelnik-Manor L, Perona P. Self-tuning spectral clustering. In: Proceedings of Advances in Neural Information Processing Systems, Vancouver, 2004. 17: 1601–1608
- 30 He J, Liu W, Chang S F. Scalable similarity search with optimized kernel hashing. In: Proceedings of International Conference on Knowledge Discovery and Data Mining, Washington, 2010. 1129–1138
- 31 Liu W, Wang J, Kumar S, et al. Hashing with graphs. In: Proceedings of International Conference on Machine Learning, Bellevue, 2011. 1–8
- 32 Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vision, 2001, 42: 145–175
- 33 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of Conference on Computer Vision and Pattern Recognition, San Diego, 2005. 886–893

# Large-scale image retrieval based on multi-feature and multi-kernel hashing learning

Xianhua ZENG<sup>1,2\*</sup>, Zhihong YUAN<sup>1,2</sup>, Guoyin WANG<sup>1,2</sup> & Jie YANG<sup>1,2</sup>

1. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and

Telecommunications, Chongqing 400065, China

\* Corresponding author. E-mail: zengxh@cqupt.edu.cn

Abstract Hashing methods can overcome the problems of low retrieval efficiency and high storage cost. Existing hashing methods use either only one feature or multiple features as the input of one kernel function. The fact that different kernel functions have different roles and different characteristics and contain different information is ignored. In this paper, an adaptive multi-feature and multi-kernel hashing learning (MFMKH) algorithm is proposed, which can adaptively combine the feature weight coefficient and the kernel weight coefficient and double-combine the multi-feature and multi-kernel advantages. The fusion of these features in the algorithm solves the disadvantage of single feature containing insufficient information. In addition, the use of a variety of different kernel functions can compensate for the lack of a single-kernel learning ability and has the dual advantages of multi-feature fusion and multi-kernel learning. Experiments on the standard datasets IRMA, Ultrasound, and Cifar10 have shown that the retrieval performance of the proposed method clearly outperforms other similar kernel-based hashing learning methods. In addition, compared to the supervised deep hashing, the retrieval performance of the proposed method is competitive on the Cifar10 dataset in the case of the reduced training time.

**Keywords** dimension reduction, multi-feature fusion, multi-kernel learning, hashing learning, adaptive learning, image retrieval



Xianhua ZENG was born in 1973. He received his Ph.D. in computer software and theory from Beijing Jiaotong University in 2009. Currently, he is a professor at Chogqing University of Posts and Telecommunications. His main research interests include image processing, machine learning and data mining.



**Zhihong YUAN** was born in 1990. Currently, he is a postgraduate at Chongqing University of Posts and Telecommunications. His current research interests include image recognition and machine learning.



**Guoyin WANG** was born in 1970. He received his Ph.D. in computer science and engineering from Xi'an Jiaotong University in 1996. Currently, he is a Changjiang scholar professor. His research interests include rough sets, granular computing, cognitive computing, intelligent information processing, data mining, and intelligent information security.



Jie YANG was born in 1987. He is a Ph.D. candidate at Chongqing University of Posts and Telecommunications. His current research interests include granular computing and machine learning.