



综述

基于深度学习的散列检索技术研究进展

袁明汶, 钱江波, 董一鸿, 陈华辉

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 大数据时代, 数据呈现维度高、数据量大和增长快等特点。面对大量的复杂数据, 如何高效地检索相似近邻数据是近似最近邻查询的研究热点。散列技术通过将数据映射为二进制码的方式, 能够显著加快相似性计算, 并在检索过程中节省存储和通信开销。近年来深度学习在提取数据特征方面表现出速度快、精度高等优异的性能, 使得基于深度学习的散列检索技术得到越来越广泛的运用。总结了深度学习散列的主要方法和前沿进展, 并对未来的研究方向展开简要探讨。

关键词: 大数据; 近似最近邻查询; 深度学习散列

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018274

Research and development of hash retrieval technology based on deep learning

YUAN Mingwen, QIAN Jiangbo, DONG Yihong, CHEN Huahui

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

Abstract: In the era of big data, data shows the characteristics of high dimension, large amount and rapid growth. How to efficiently retrieve similar data from a large amount of complex data is a research hotspot. By mapping data to binary codes, the hashing technique can significantly accelerate the similarity calculation and reduce storage and communication overhead during the retrieval process. In recent years, deep learning has shown excellent performance in extracting data features. Deep learning-based hash retrieval technique has the advantages of high speed and accuracy. The methods and advanced development of deep learning hashing were mainly summarized, and the future of research direction was briefly discussed.

Key words: big data, approximate nearest neighbor query, deep learning hashing

1 引言

近年来, 信息技术的快速发展使得互联网、

物联网、新闻媒体以及其他行业的数据呈爆炸式增长, 人们进入了一个大数据时代。大数据^[1]时代, 数据呈现出量大、维度高、种类多样、价值

收稿日期: 2018-03-18; 修回日期: 2018-10-10

通信作者: 钱江波, qianjiangbo@nbu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61472194, No.61572266); 浙江省自然科学基金资助项目 (No.LY16F020003)

Foundations Items: The National Natural Science Foundation of China (No.61472194, No.61572266), Zhejiang Provincial Natural Science Foundation of China (No.LY16F020003)

稀疏和时效要求高等特点。大数据在各个领域得到了广泛运用^[2-3],已成为一种重要的战略资源,同时也对数据的存储、管理和分析提出了新的挑战,这些问题成为了学术界和工业界关注的热点。处理和分析大数据是为了有效利用大数据资源,因此如何对海量的数据快速检索是大数据研究的主要内容之一。

传统的检索方法主要分为两类。一类是线性扫描的检索方式,准确率高,但是处理高维数据计算量较大,检索效率较低;另一类是基于树的检索技术,即创建树形索引结构,索引空间按照不同条件划分成不同的子空间,算法检索效率一般为 $O(\log n)$,低于暴力搜索的效率 $O(n)$ 。但是随着数据维数的增加,存储树结构也会带来很大的存储和更新开销,使得基于空间树的索引结构在搜索效率上急剧下降。

由于数据的维度灾难,高维数据的精确搜索计算时间长、存储开销大,与线性扫描的检索方式相比没有明显的优势,而由精确解转向近似解不失为一种有效的解决方法,因此基于散列的近似最近邻(approximate nearest neighbor, ANN)搜索技术^[4-9]引起了国内外研究人员的广泛关注,在实践中发现该方法具有很好的近似检索效果,相比于精确搜索在检索效率上有很大数量级的提高。

散列技术的基本思想是构建一系列散列函数,将高维空间中的数据对象映射为二进制特征向量,以使相似的数据对象被映射成相似的二进制代码。通过将高维实体特征向量编码为低维紧凑二进制码,可以显著加快相似性计算,并在搜索过程中节省存储空间。基于散列的检索方法主要分为两类:数据独立^[10-14]和数据依赖^[4,15-21]。数据独立的散列方法是采用随机投影将高维空间中的数据映射到低维空间,然后执行二值化计算得到二进制码,其中具有代表性的方法有局部敏感散列(locality sensitive hashing, LSH)^[10]以及基

于核的^[14,21-23]和区分性扩展^[13,17]的局部敏感散列。这些散列方法中,LSH只有较长的散列码才能达到适用的性能,其他散列方法也不能有效地利用样本的几何和判别信息构造散列函数,这造成实际使用中效率较低。因此,依赖于数据的散列方法是更可取的,许多这样的算法最近已被提出。数据依赖的散列方法是使用各种统计学习技术学习散列函数,以将数据映射成二进制码的形式。已有的先进方法包括谱散列^[9]、迭代量化(iterative quantization, ITQ)^[4]、K-means散列^[16]、最小损失散列(minimal loss hashing, MLH)^[19]等。这些基于学习的散列方法在许多应用中取得了相当好的效果,但它们大多为简单的线性投影,在捕捉数据的非线性结构方面还有很大缺陷。

近年来,人工智能技术在解决许多问题中取得了突破性的发展,驱动人工智能发展的深度学习^[24]技术受到了越来越多的关注。深度学习广泛应用于计算机视觉、机器翻译、语音识别以及文本挖掘等多个领域,其高水平的数据表征能力,能够提取通用性高和顽健性好的表征。

深度散列就是利用深度学习学习方法学习二进制散列编码,因此深度学习的散列检索技术表现出较高的性能。基于深度学习的散列方法不断涌现出新的研究成果,使得深度散列成为了目前的研究热点。

2 深度学习散列方法概述

深度学习是直接对原始数据做非线性转换,提取通用性高、顽健性好的高层数据表征表示属性类别。近年来提出了各种各样的深度学习算法^[24],在各种机器学习任务中取得了较好的效果,包括基于深度学习的散列技术。本文根据深度学习散列网络结构的发展不同,把深度学习散列分为4个主要类别:首次使用深度学习的散列技术是2009年由Hinton提出的基于受限玻尔兹曼机(restricted Boltzmann machine, RBM)的深度散列方法;反



向传播 (back propagation, BP) 神经网络的广泛运用, 2012 年开始出现了基于多层感知机的深度散列方法; AlexNet^[25]在 ImageNet 图像识别任务中取得冠军, 深度学习方法倍受关注, 2014 年卷积神经网络散列 (convolutional neural network hashing, CNNH)^[33]出现后, 基于卷积神经网络 (convolutional neural network, CNN) 的深度散列方法不断涌现; 2014 年提出的生成对抗网络 (generative adversarial network, GAN) 已经在各种图像生成中被大量使用, 近年来也出现了基于 GAN 的散列学习方法。深度散列技术发展导图如图 1 所示。

3 基于受限玻尔兹曼机的深度散列方法

受限玻尔兹曼机是玻尔兹曼机的一种特殊拓扑结构, 是 1986 年提出的一种生成式随机神经网络。这种网络中, 当神经元为激活状态时, 神经元的输出用 0 或 1 表示; 当神经元为未激活状态时, 神经元的输出用 1 或 0 表示, 其中的神经元都为随机神经元。

3.1 语义散列

语义散列 (semantic hashing, SH)^[26]是最早

提出的基于深度学习的散列方法。该方法首先构造一个深度生成模型, 学习数据隐藏的二进制单元 (比如: 潜在的主题特征), 这种隐藏的二进制单元能够模拟输入的文本数据 (比如: 单词计数向量)。该深度模型可被视作由若干受限玻尔兹曼机^[27]堆叠而成, 采用预训练以及微调两步操作训练。深度模型由多个 RBM 堆叠而成, 预训练阶段需要自下向上逐层训练每一个 RBM, 训练下一层 RBM 模型得到的输出作为上一层的数据输入, 最后一层的训练输出位数为最终散列编码位数, 从而得到了一个由多层 RBM 堆叠而成的深度散列模型。最后微调阶段通过联合层与层之间的关系对每一层的 RBM 做微调修正, 从而得到自动编码器。训练后, 得到一个多层的受限玻尔兹曼机模型, 对于任何一个数据而言, 只需要简单地对最后一层的输出进行二值化处理, 即可得到该数据的散列编码。实验表明, 由这种深度 RBM 方法得到的散列编码能够将输入文档之间的语义相似关系保持下来, 其中每一个散列编码被作为一个存储地址来定位相应的数据位置。这种方法中, 本身语义相似的数据会被映射至相邻的存储地址中, 因此能够通过散列查找实现高效的数据检索。

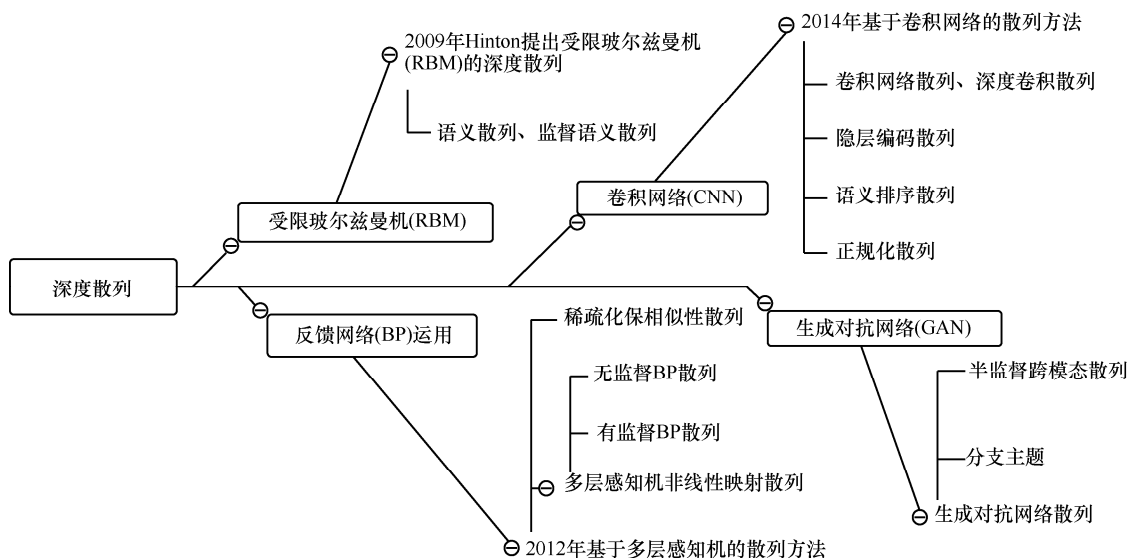


图 1 深度散列技术发展导图

上述语义散列深度模型虽然具有一定的非线性表征能力,但是需要对原始数据手工提取特征后输入网络,散列性能在很大程度上依赖于特征提取的效果,模型每层的微调修正在训练大量数据中也较为困难,与普遍意义上的深度学习算法还有很大差别,因此基于语义散列出现了一些改进。

3.2 监督型语义散列

为了使语义散列的效果得到进一步提升,通过借鉴非线性近邻成分分析(neighbourhood components analysis, NCA)^[28]嵌入的思想,提出了一种名为监督型语义散列(semi-supervised hashing, SSH)^[29]的散列方法。该方法中,给定训练样本之间的近邻关系作为监督信息,在最后一层上对目标函数进行优化,生成有指导作用的散列编码。有监督的深度RBM方法,由于第一层采用高斯分布可视化单元建模,因此应对大规模图像数据检索效果非常显著,使得该方法可以被应用到除文本数据以外的其他数据集中。

4 基于BP神经网络的深度散列方法

BP神经网络是一种误差反向传播网络,是目前使用最为广泛的多层前馈神经网络。网络的训练分为前馈计算和反向误差修正计算,前馈计算即上一层的输出作为下一层的输入,直到计算至最后一层,如果最后一层输出不能得到期望输出,则进行反向误差修正计算,通过梯度下降的方式将误差损失沿连接通路反向传导,更新权值参数,前馈计算和反馈计算多次迭代,直到收敛至满意的输出为止。

4.1 稀疏化保相似性散列

稀疏化保相似性散列(sparse similarity-preserving hashing, SSPH)^[30]是针对已有散列方法的性能和计算代价难以两全的问题,试图以增加散列码长度为代价来提升散列性能和降低计算开销的方法,这种方法对先前的大部分散列技术都造成了一定的影响。稀疏化保相似性散列是

有监督的散列学习方法,监督信息由原始数据两两之间的相似性提供,利用监督信息通过训练定制的前馈神经网络得到其散列函数族,即一组非线性嵌入函数。模型结构中,训练阶段有两个进行快速相似性稀疏化编码的ISTA—类型网络^[31],这两个网络耦合并且共享一组相同的参数。每个网络中的输出向量稀疏且连续,利用双曲正切函数对得到的输出向量进行二值化操作,得到最后的二进制散列编码。

4.2 基于多层感知机无监督的深度散列

多层感知机深度散列^[32]首先是构建一个深度多层感知模型,通过此模型将图像数据映射成紧凑的实值向量,再经过二值化操作生成紧凑二进制散列码,然后利用该二进制编码来构建散列索引结构,实现大规模图像检索。该深度散列模型在构建的深度散列网络顶层设计了3个约束条件:最小化图像的原始实值特征向量以及图像二进制编码之间的重构误差;二进制码在每个比特位上均匀分布;不同的比特位尽可能独立,即各投影维度间正交不相关。这种约束方法在一些传统的无监督散列方法中也被广泛使用,如迭代量化^[4]。

该深度散列的主要思想是最小化量化损失,提升散列码的性能,制定目标约束学习深度散列模型中的网络参数,设计的最小化损失函数为:

$$\begin{aligned} \min_{w,c} J &= J_1 - \lambda_1 J_2 + \lambda_2 J_3 + \lambda_3 J_4 \\ J_1 &= \frac{1}{2} \|B - H^M\|_F^2 \\ J_2 &= \frac{1}{2N} \text{tr}(H^M (H^M)^T) \\ J_3 &= \frac{1}{2} \sum_{m=1}^M \|W^m (W^m)^T - I\|_F^2 \\ J_4 &= \frac{1}{2} \sum_{m=1}^M (\|W^m\|_F^2 + \|c^m\|_2^2) \end{aligned} \quad (1)$$

其中, J_1 为最小化学习的二进制向量与紧凑的实值向量之间的编码损失,使得学习的紧凑特征可以在二进制码中很好地保持; J_2 为最大化学习的



二进制向量的方差, 以确保各比特均匀分布, 迫使散列特征尽可能紧凑地分布; J_3 为对投影矩阵执行松弛的正交约束, 从而使每个变换的独立性最大化; J_4 为控制参数大小的正则项; λ_1 、 λ_2 、 λ_3 3 个参数用于平衡不同损失项对总量化损失的影响。

无监督的多层感知机散列, 对散列网络的投影规则限制约束, 能在一定程度上提升散列性能, 但是无法显著地发掘数据间的相似性关系, 所以散列性能并不高。

4.3 基于多层感知机有监督的深度散列

有监督的深度散列模型和无监督的深度散列模型的主要区别在于利用了标签中的监督信息来构造散列方法的目标函数。在实验中可以证明与无监督的深度散列相比, 由于其使用了指导信息生成的散列码表现出更高的性能。深度散列无监督方法 (deep hashing, DH)^[32] 和有监督方法 (supervised deep hashing, SDH)^[32] 相比, 其性能同比提升近一倍, 但是相比现有的端到端的模型, 其性能还有待进一步提高。

以上方法中有一个共同问题, 即在对深度神经网络进行训练的过程中没有预训练阶段, 散列编码直接从一组训练数据中学习得到。然而, 缺少预训练过程可能会导致学习生成的二进制编码表现欠佳。特别地, 稀疏化保相似性散列在某些图像数据集上的查询精度要低于顶尖的基于核函数的有监督散列学习方法 (KSH)^[33]; 深度散列以及有监督的深度散列的性能分别优于 ITQ 以及 ITQ 的有监督版本 CCA+ITQ^[4,32]。

5 基于深度卷积神经网络的深度散列方法

自 2012 年卷积网络 AlexNet 模型在 ImageNet 图像识别任务中获得冠军之后, 卷积网络在深度学习方法中倍受关注, 卷积神经网络在众多网络中得到了最深入的研究, 基于卷积神经网络的深度学习散列方法不断涌现。

5.1 深度卷积神经网络散列

卷积神经网络具有自动提取图像高维特征的能力, 可以显著提高图像识别任务的效果。受到卷积神经网络强大的数据表征能力的影响, 提出的深度卷积神经网络散列^[34]方法, 对散列学习的研究者们产生了重大的影响。CNNH 把散列学习分为两个阶段: 首先学习相似性散列码, 给定 n 个图片 $I = \{I_1, I_2, \dots, I_n\}$ 和成对的相似矩阵 S (如果 I_i 和 I_j 语义相似 $S_{ij}=1$, 不相似则 $S_{ij}=-1$), 通过构建相似度矩阵, 学习近似散列码最小化重构错误, 即让散列码内积矩阵与相似度矩阵之差最小, 如式 (2) 所示:

$$\min_H \sum_{i=1}^n \sum_{j=1}^n \left(S_{i,j} - \frac{1}{q} H_i \cdot H_j^T \right)^2 = \min_H \left\| S - \frac{1}{q} H H^T \right\|_F^2 \quad (2)$$

其中, $\|\cdot\|$ 是 F 范数, $H \in \{-1, 1\}^{n \times q}$ 。训练图像 H 编码的近似散列码保持了 S 中成对的相似性, 由于 H 的离散性很难直接优化, 因此松弛处理为 $H \in [-1, 1]^{n \times q}$, 采用牛顿方向的坐标下降算法求解。

在得到散列码后, 第二个阶段提取图像特征和学习散列函数同时进行, 构建散列模型的过程中使用三层的深度卷积网络学习图像特征, 并在卷积网络的末端添加一个全连接层和散列层, 网络的训练可以看作对一个多标签分类问题的求解, 通过最小化交叉熵损失来达到降低散列层的实际输出值和离散的二进制散列码之间的损失。此外, 本文还通过加入分类的交叉熵损失 (CNNH+) 提供更多的监督信息来提升散列性能。

CNNH 模型与传统的手工提取特征的散列方法相比, 虽然在散列检索的性能上得到了显著的提升, 但训练阶段散列码的生成不是由网络直接学习而来的, 因而学习得到的图像表征不能反作用于散列码的生成, 没有充分发挥深度学习的优势。因此, 后面又提出了许多改进方法。

5.2 深度神经网络散列

深度神经网络散列 (network in network hash-

ing, NINH)^[35]方法相比已有的深度散列方法实现了端到端的散列学习,能够同时学习得到图像的高层特征以及对应的散列值。网络使用三元组图像训练,用 I 表示图像数据集,则由三张图像构成的三元组表示为 (I, I^+, I^-) , 其中 I 相比 I^+ 更相似于 I^+ , 这种监督学习方法如今在深度学习技术中使用越来越广泛。散列学习的目标是找到一组函数映射 $F(\cdot)$ 学习二进制码,即二进制码 $F(I)$ 更接近于 $F(I^+)$ (相比 $F(I^-)$)。因此,三元组损失函数的设计方案是:在生成的二进制码中,相似样本二进制的海明距离尽可能小,不相似样本的海明距离尽可能大。为了达到这个目标,损失函数设计为:

$$l_{\text{triplet}}(F(I^+), F(I^-)) = \max(0, 1 - (\|F(I) - F(I^-)\|_H - \|F(I) - F(I^+)\|_H))$$

$$\text{s.t. } F(I), F(I^+), F(I^-) \in \{0, 1\}^q \quad (3)$$

其中, $\|\cdot\|_H$ 表示海明距离。式 (3) 为了便于求导优化,采用 l_2 范数代替海明范数和范围约束替代整数约束。

此外为了学习优质的散列码,笔者对散列网络做了特殊的设计,主要包括以下两个方面。

(1) 对全连接层进行划分,使用部分连接层代替全连接层,每个部分只学习一个比特位,减少了比特位之间的冗余性,如图 1(a)所示。

(2) 为了促使分割和编码模块的输出是二进制编码,使用 sigmoid 激活函数做离散松弛约束,利用分段量化方案来生成一个近似的离散散列编码。

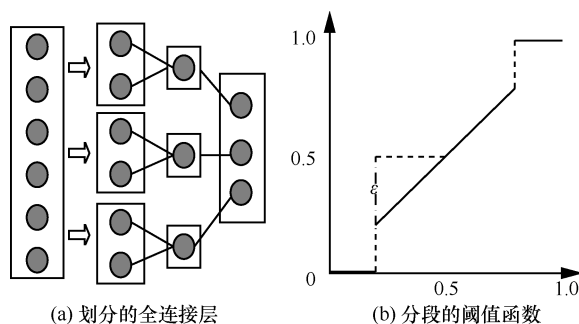


图2 NINH 网络结构和阈值函数

对于输入变量 $s = \text{sigmoid}(c) \in [0, 1]$, 分段量化方案如图 1(b)所示,这个分段量化函数设计为:

$$g(s) = \begin{cases} 0, & s < 0.5 - \varepsilon \\ s, & 0.5 - \varepsilon \leq s \leq 0.5 + \varepsilon \\ 1, & s > 0.5 + \varepsilon \end{cases} \quad (4)$$

其中, ε 是一个小于 1 的正超参数,当 sigmoid 函数输出在 $[0, 0.5 - \varepsilon)$ 和 $(0.5 + \varepsilon, 1)$, 它们分别被截断为 0 或 1。

三元组输入和有针对性散列编码网络两部分合在一起,构成了一个端到端的散列模型,模型学到的图像表征能够使散列码也得到更新,由此使得图像特征学习过程以及散列值学习过程紧紧耦合、相辅相成,因此这种方法的性能更优于卷积神经网络散列。

5.3 基于语义排序的多标签图像检索

DSRH (deep semantic ranking hashing)^[36]是一种基于标签学习保语义相似性的散列方法,同样也是一种基于 CNN 的散列方法。笔者认为大多数散列方法只能对数据做简单的二进制相似性编码,对于复杂的多标签相关联的图像数据无法有效地编码。因此,提出了一种基于深层语义排序的散列函数学习算法,以保持多标签图像之间的多级语义相似性。本文将深度卷积神经网络运用到构造散列函数中,同时学习数据的特征表示和散列函数,避免了手工特征的语义表征能力的局限性。同时,采用编码多级相似性信息的排序列表来指导这种深度散列函数的学习,即直接让网络学习的排序列表,对最终的评测指标优化。直接优化这样的排名标准非常困难,需要解决学习过程中涉及的非平滑和多变量排序导致的难以解决的优化问题。因此,笔者定义了一个凸上界排序损失函数作为替代,如式 (5) 所示:

$$L(h(q), \{h(x_i)\}_{i=1}^M) = \sum_{i=1}^M \sum_{j: r_j < r_i} w(r_i, r_j) [\delta d_H(h(q), h(x_i), h(x_j)) + \rho]_+ \quad (5)$$



其中, M 是排名列表的长度, $[\cdot]_+ = \max(0, \cdot)$, $\delta d_H(h, h_1, h_2) = d_H(h, h_1) - d_H(h, h_2)$, $d_H(\cdot, \cdot)$ 是海明距离, ρ 是一个范围参数, 它控制数据对之间距离的最小范围。

为了让排名靠前的检索结果有一个更大的增益因子得分, 式 (5) 权重 w 设计为:

$$w(r_i, r_j) = \frac{2^{r_i} - 2^{r_j}}{Z} \quad (6)$$

其中, Z 是归一化常数, r_i 是在排行榜的第 i 个数数据库点的相似程度。 x_i 和 q 的相关性高于 x_j 和 q 的相关性, 如果 x_i 排在 x_j 之后, 则得分的下降幅度越大。

深度语义排序散列利用一组三元组上的排序损失作为替代损失, 解决了非光滑和多变量排序衡量的优化问题, 使得随机梯度下降可以有效地优化模型参数。由于没有在网络特征提取能力和减少散列码量化损失上做显著改善, 而直接对评测指标进行优化, 导致实际使用中检索效果并不好。

5.4 正则化深度散列

深度正则化相似性比较散列 (deep regularized similarity comparison hashing, DRSCH)^[37] 同样是一个监督学习的散列框架, 能够直接从原始图像生成紧凑和可扩展的散列码。笔者认为大多数以前的方法都预先设置了输出散列码的比特长度, 忽略了不同比特的显著性水平, 限制了其实际的灵活性。因此, 该算法也是基于三元组构造散列损失, 提出了加权海明距离的思想, 设计加权海明距离计算式为:

$$H(h(x_j), h(x_k)) = h(x_j) \tilde{w} h(x_k) = - \sum_i w_i^2 h_i(x_j) h_i(x_k) \quad (7)$$

其中, $\tilde{w}$ 是对角矩阵, $\tilde{w}(i, j) = w_i^2$, 权值作为网络参数在训练过程中更新, 通过更新网络得到合适的权值参数。该加权海明距离虽在时间复杂度上有所上升, 但是海明距离计算更为精确, 同时

能够通过权值控制散列码生成的长度。

另一方面, 使用正则项控制图像之间的邻接关系, 提升散列编码效果。设计正则函数为:

$$\sum_{i,j} \psi_w(h_i, h_j) = \frac{1}{2} \text{tr}(HLH^T) \quad (8)$$

其中, $H = \begin{bmatrix} h_1 \tilde{w}^{\frac{1}{2}}, h_1 \tilde{w}^{\frac{1}{2}}, \dots, h_M \tilde{w}^{\frac{1}{2}} \end{bmatrix}$, M 为图像的个数, $L = U - S$, S 为相似度矩阵, U 为对角矩阵, $U_{ii} = \sum_j S_{i,j}$ 。因此, 整体的三元组最小化损失函数设计为:

$$L_{\text{loss}} = \min_{w, \omega} \sum_{i=1}^N \max \{G_w(h_i, h_i^+, h_i^-), C\} + \lambda \text{tr}(HLH^T) \quad (9)$$

其中, $G_w(h_i, h_i^+, h_i^-) = H(h_i, h_i^+) - H(h_i, h_i^-)$, C 为常数, λ 为超参数。本文在深度散列学习框架中引入了正则化相似度比较公式, 以保证图像邻接一致性, 使用加权的方式权衡散列码, 从而可以方便地获得比特的可伸缩性。因此, 不仅在检索准确性方面得到提升, 而且提高了不同长度散列的灵活性。

在深度监督散列 (deep supervised hashing, DSH)^[38] 中同样也采用了正则项的方式, 约束散列层的实值输出范围, 使其与散列码相接近。网络学习的目标是使学习得到的散列码相似图片尽可能相似, 不相似的尽可能不相似。因此, 设计最小化损失函数为:

$$L = \sum_{i=1}^N \left\{ \frac{1}{2} (1 - y_i) D_h(b_{i,1}, b_{i,2}) + \frac{1}{2} y_i \max(m - D_h(b_{i,1}, b_{i,2}), 0) \right\} \quad (10)$$

其中, $b_{i,j} \in \{+1, -1\}^k$, $i \in \{1, \dots, N\}$, $j \in \{1, 2\}$ 。 $D_h(\cdot, \cdot)$ 表示两个二进制向量之间的海明距离, $m > 0$ 是一个范围阈值参数。由于 $b_{i,j}$ 获取需要对网络输出进行阈值化处理, 如果完全忽略二进制约束, 则欧氏空间和海明空间之间的差异, 会导致次优的二进制码。因此, 采用正则化的方式做二元约束, 设计松弛的损失函数为:

$$L_r = \sum_{i=1}^N \left\{ \frac{1}{2} (1 - y_i) \|b_{i,1} - b_{i,2}\|_2^2 + \frac{1}{2} y_i \max(m - \|b_{i,1} - b_{i,2}\|_2^2, 0) + \alpha (\|b_{i,1}\|_1 - 1 + \|b_{i,2}\|_1 - 1) \right\} \quad (11)$$

DSH 通过深度网络提取图像的非线性特征, 使用正则项约束减少实值网络输出空间和所需的海明空间之间的差异化, 同时利用图片对的监督信息生成期望的海明空间, 相比传统的散列方法能够更快地对新的图像进行编码, 是一种简单有效的散列方法。

5.5 隐藏层编码的深度散列

二进制编码深度散列 (deep learning of binary hash codes, DLBHC) [39] 是直接通过隐藏层表示散列码的有监督散列方法, 即在已有的网络框架的全连接层的倒数第一、第二层之间添加一个散列层, 并用散列层的输出学习图像的分类信息, 构建分类损失学习整个网络参数。通过对散列层输出 (Out(H)) 阈值分割, 从而得到二进制散列码:

$$H^j = \begin{cases} 1, & \text{Out}^j(H) \geq 0.5 \\ 0, & \text{其他} \end{cases} \quad (12)$$

让 $\Gamma = \{I_1, I_2, \dots, I_n\}$ 表示用于检索的 n 个图像组成的数据集, 则每个图像对应的散列码被表示为 $\Gamma_H = \{H_1, H_2, \dots, H_n\}$, 其中 $H_i \in \{0, 1\}^h$ 。该散列框架主要包括 3 个步骤: 用 ImageNet 数据集对卷积网络进行预训练, 学习图像的中层特征表示; 添加隐藏层, 在对应的数据集上微调模型, 在这一层神经元上学习散列表示; 利用散列码对查询图像进行检索。

监督语义保持的深度散列 (supervised semantics-preserving deep hashing, SSDH) [40] 也采用在分类层前面添加散列层的网络结构, 通过散列值重构分类信息。笔者在从两个方面分别对网络的损失函数做了改进, 首先分类误差采用改进的交叉熵损失, 计算式如下:

$$\arg \min_w E_1(W) = \arg \min_w \sum_{n=1}^N L(y_n, \hat{y}_n) + \lambda \|W\|^2 \quad (13)$$

其中, $L(\cdot)$ 是最大限度减少分类误差的损失函数, W 表示网络的权重, λ 是控制正则项的相对重要性参数。改进后的分类损失函数 $L(y_n, \hat{y}_n)$ 为:

$$L(y_n, \hat{y}_n) = \sum_{m=1}^M \begin{cases} 0, & y_{nm} = 1 \text{ \& } \hat{y}_{nm} \geq 1 \\ 0, & y_{nm} = 1 \text{ \& } \hat{y}_{nm} \geq 0 \\ \frac{1}{2} |y_{nm} - \hat{y}_{nm}|_p^p, & \text{其他} \end{cases} \quad (14)$$

其次, 笔者在量化误差上对函数做了进一步改进。让大于 0.5 的输出值尽可能地接近 1, 小于 0.5 的输出值尽可能地接近 0; 同时保证各个比特位均匀分布, 优化以下损失函数获得二进制码:

$$\arg \min_w -\frac{1}{k} \sum_{n=1}^N \|a_n^H - 0.5e\|_p^p + \sum_{n=1}^N |\text{mean}(a_n^H) - 0.5e|^p = \arg \min_w -E_2(W) = E_3(W) \quad (15)$$

其中, $p \in \{1, 2\}$, a_n^H 表示散列层输出值, K 表示散列码长度, $E_2(W)$ 表示迫使散列层输出值接近于 0 或 1, $E_3(W)$ 表示进一步确保每个节点的输出接近 50% 的概率为 0 或 1。构造相似性保持的整个目标函数为:

$$\arg \min_w = \alpha E_1(W) - \beta E_2(W) + \gamma E_3(W) \quad (16)$$

其中, α 、 β 、 γ 是每个项的权重。

SSDH 在分类损失和量化损失两个方面做了很大改进, 因此相比 DLBHC 方法在检索效果上得到了很大提升。两个隐藏层散列方法都不依赖于相似性数据对, 适合在大数据集上使用。散列语义重构的方法虽然在生成的散列码上保留了语义信息, 但设计的损失函数没有考虑数据点之间的相对距离, 无法保证生成散列码的保相似性。

5.6 半监督的跨模态 GAN 散列

生成式对抗网络是 2014 年由 Goodfellow 等人提出的一种深度学习网络。模型主要由生成模



型和判别模型两个模块组成, 生成模型 G 捕获数据的分布, 判别模型 D 估计来自训练数据的概率, G 和 D 在相互博弈中学习好的输出。

有监督的跨模态散列方法 (semi-supervised cross-modal hashing by generative adversarial network, SCH-GAN) [41] 通过结合语义信息取得了相当大的进展。然而, 它们主要有两个局限性: 由于涉及多种模式, 严重依赖手工标记且难以获得大规模标签的跨模态训练数据; 忽略了不同模态大量未标记数据中包含的丰富信息, 特别是来自另一种模式的容易被错误检索的边缘示例, 这可以帮助模拟不同模态之间的相关性。针对此问题笔者提出了一个新的半监督的生成对抗网络跨模态散列算法, 大量的学习无标签数据中包含的丰富信息。生成模型选择信息未标记的数据与给定的标记查询形成数据对来欺骗判别模型, 而判别模型试图区分是生成的数据对还是真的数据对。这两个模型扮演两个玩家的角色, 扮演一个极小游戏来优化对方, 提升跨模态散列性能。对抗性过程的损失定义如下:

$$V(G, D) = \min_{\theta} \max_{\phi} \sum_{j=1}^n (E_{i \sim p_{\text{true}}(i^L | q_i^j, r)} [\log(D(i^L | q_i^j))] + (E_{i \sim p_{\theta}(i^U | q_i^j, r)} [\log(1 - D(i^U | q_i^j))]) \quad (17)$$

式(17)用于文本查询图像任务, 而图像查询文本任务类似地定义为:

$$V(G, D) = \min_{\theta} \max_{\phi} \sum_{j=1}^n (E_{i \sim p_{\text{true}}(i^L | q_i^j, r)} [\log(D(t^L | q_i^j))] + (E_{i \sim p_{\theta}(i^U | q_i^j, r)} [\log(1 - D(t^U | q_i^j))]) \quad (18)$$

式(17)和式(18)是对称的, 因此下面仅使用文本查询图像任务呈现详细的目标函数。生成模型 G 被表示为 $p_{\theta}(i^U | q_i^j, r)$, 定义 Softmax 函数为:

$$p_{\theta}(i^U | q_i^j, r) = \frac{\exp(-\|h_T(q_i) - h_I(i^U)\|^2)}{\sum_{i^U} \exp(-\|h_T(q_i) - h_I(i^U)\|^2)} \quad (19)$$

其中, $h_i(*)$ 和 $h_t(*)$ 分别表示图像和文本的散列函数。判别模型 D 预测选择的图像与文本相关的概率, 定义 sigmoid 相关评分为:

$$D(i^U | q_i) = \text{sigmoid}(f_{\phi}(i^U, q_i)) = \frac{\exp(f_{\phi}(i^U, q_i))}{1 + \exp(f_{\phi}(i^U, q_i))} \quad (20)$$

$$D(i^L | q_i) = \text{sigmoid}(f_{\phi}(i^L, q_i)) = \frac{\exp(f_{\phi}(i^L, q_i))}{1 + \exp(f_{\phi}(i^L, q_i))} \quad (21)$$

相关评分 $f_{\phi}(i^U, q_i)$ 和 $f_{\phi}(i^L, q_i)$ 三元组排序损失定义如下:

$$f_{\phi}(i^U, q_i) = \max(0, m_i + \|h_T(q_i) - h_I(i^+)\|^2 - \|h_T(q_i) - h_I(i^U)\|^2) \quad (22)$$

$$f_{\phi}(i^L, q_i) = \max(0, m_i + \|h_T(q_i) - h_I(i^L)\|^2 - \|h_T(q_i) - h_I(i^-)\|^2) \quad (23)$$

其中, i^+ 表示图片与查询文本 q_i 语义相似, i^- 表示语义不相似的图片查询样例, i^U 表示通过生成模型选择的图片, m_i 为超参数。

SCH-GAN 散列模型采用半监督的方法, 同时利用有标签的数据和无标签数据训练模型, 使大量数据无标签的数据利用起来, 在大规模数据检索中无标签的数据占据主要部分, 因此该模型更加符合真实的使用场景, 同时其多模态查询的特性也给实际查询提供了更多的检索方案。

6 性能分析

早期的深度散列方法 (如 SH、SSPH) 有一定的非线性学习能力, 但网络无法输入原始数据, 需要对原始数据手工提取特征, 在性能方面与传统的散列方法相比提升不大, 但训练复杂度较高。近期, 深度学习的进一步发展, 可以直接通过原始数据学习高层特征表示来提升学习性能。

CNNH 采用卷积网络直接从原始数据学习高层特征表示, 很大程度地提升了散列性能, 但该模型没有实现端到端的学习, 学习到的高层表征不能反作用于散列码的学习。NINH 提出了一种新的网络结构, 使用三元组数据训练网络, 实现了端到端的学习, 进一步提升了散列学习的性能, 但量化损失的设计还不够细致。DSRH 让网络直接学习这个排序, 即直接优化评测指标, 但直接对排序指标优化的方式并不容易, 实际效果提升有限。DRSCH 和 DSH 使用正则项约束图像之间的近邻关系和实值的输出空间, 也有效地提升了散列性能, 但是没有充分利用标签的语义信息。DLBHC、SSDH 提出散列层输出重构语义标签信息的思想, 使得散列性能得到了较大的提升, SSDH 在分类损失和量化损失上做了细致的改进使得散列性能达到一个新的高度。虽然基于散列语义重构的思想较大地提升了散列检索的性能, 但设计的损失函数没有考虑数据点之间的相对距离, 生成散列码的保相似性无法保证, 因此会产生检索任务的稳定性问题。GAN 散列由于其无监督的特性, 能够使用大量的无标签数据训练模型, 但是模型的训练相比传统的深度模型更为困难, 训练过程中需要保证样本数据的多样性分布。散列检索运用范围很广, 紧凑的保相似性散列码可以显著加快相似度计算, 并在搜索过程中节省存储空间, 但散列搜索的准确度是一个关键问题。相比于传统的散列方法, 大部分深度散列方法都是在提升准确性上进行改进, 如设计新的网络结构学习更优

图像表示、不同的损失函数和替代的激活函数减少量化损失。

本文将近几年深度散列的性能提升和采用的方法做了详细比较, 见表 1。

其中, SCH-GAN 为跨模态散列, 模型采用广泛使用的跨模态检索数据集 WIKIPEDIA 训练评测, 其由 2 866 对图像/文本组成, 表格实验结果对应: 图像查文本/文本查图像。

7 结束语

散列^[42]技术将原始数据映射为保相似性二进制码, 有效地提高了相似性计算速度, 能够显著减少数据的存储和通信开销, 深度散列利用深度学习思想使散列技术得到了进一步发展。深度散列模型具有许多优点: 深度散列将原始数据直接输入网络, 做非线性变换, 得到数据的高层特征表示, 进而得到可靠的散列函数, 实现有效的散列查询; 由于深度散列具有可迁移性, 面对新数据只需微调模型参数, 具有可迁移性; 深度散列直接从原始数据学习散列函数, 可以有效避免手工提取特征带来的误差, 得到的散列函数具有更好的顽健性。

虽然把深度学习运用到散列技术中取得了显著的效果, 但是仍有许多方面需要进一步研究: 由于 sigmoid/tanh 激活函数的饱和性质, 当输出接近期望值 (0/1 或 -1/+1) 时梯度变小, 网络训练困难, 寻找有效的替代品, 可能是一个不错的解决方案; 散列检索速度是一个关键指标, 散列码生

表 1 深度散列在 CIFAR10 上性能比较

算法	12 bit(mAP)	32 bit(mAP)	48 bit(mAP)	采用方法
CNNH	0.465	0.521	0.532	散列码学习、学习散列函数
NINH	0.552	0.558	0.581	网络分割、阈值损失
DRSCH	—	0.628	0.630	海明码加权、正则约束
DSH	0.615	—	0.675	正则约束、保相似性损失
SSDH	0.872	0.911	0.908	散列重构、编码约束
SCHGAN	0.523/0.859	0.530/0.876	0.548/0.883	半监督的生成对抗学习



成需要经过多层网络的前馈传播计算, 散列编码生成所需要的时间也是一个值得关注的问题; 现有的深度散列方法主要运用于图像的检索, 如何将深度散列方法广泛运用于各个方面, 解决一些新的问题, 还需要展开进一步探索。

参考文献:

- [1] 吴军. 大数据和机器智能对未来社会的影响[J]. 电信科学, 2015, 31(2): 7-16.
WU J. Big data, machine intelligence and their impacts to the future world [J]. Telecommunications Science, 2015, 31(2): 7-16.
- [2] 尤海浪, 钱锋, 黄祥为, 等. 基于大数据挖掘构建游戏平台个性化推荐系统的研究与实践[J]. 电信科学, 2014, 30(10): 27-32.
YOU H L, QIAN F, HUANG X W, et al. Research and practice of building a personalized recommendation system for mobile game platform based on big data mining [J]. Telecommunications Science, 2014, 30(10): 27-32.
- [3] 徐雅斌, 刘超, 武装, 等. 基于用户兴趣和推荐信任域的微博推荐[J]. 电信科学, 2015, 31(1): 13-20.
XU Y B, LIU C, WU Z, et al. Micro-blog recommendation based on user interests and recommendation trust domain [J]. Telecommunications Science, 2015, 31(1): 13-20.
- [4] GONG Y, LAZEBNIK S. Iterative quantization: a procrustean approach to learning binary codes[C]// IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Colorado Springs, CO, USA. Piscataway: IEEE Press, 2011: 817-824.
- [5] HE J, LIU W, CHANG S F. Scalable similarity search with optimized kernel hashing[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 25-28, 2010, Washington, DC, USA. New York: ACM Press, 2010: 1129-1138.
- [6] LIU X, HE J, LANG B, ET AL. Hash bit selection: a unified solution for selection problems in hashing[C]// Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. Piscataway: IEEE Press, 2013: 1570-1577.
- [7] WANG J, KUMAR S, CHANG S F. Semi-supervised hashing for large-scale search[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(12): 2393-2406.
- [8] WANG J, SHEN H T, SONG J, et al. Hashing for similarity search: a survey[J]. Computer Science, 2014.
- [9] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing[C]// International Conference on Neural Information Processing Systems, Dec 8-10, 2008, Vancouver, British Columbia, Canada. New York: ACM Press, 2008: 1753-1760.
- [10] ANDONI A, INDYK P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions[C]// 2006 IEEE Symposium on Foundations of Computer Science, Oct 21-24, 2006, Berkeley, CA, USA. Piscataway: IEEE Press, 2006: 459-468.
- [11] CHUM O, PHILBIN J, ZISSERMAN A. Near duplicate image detection: min-hash and tf-idf weighting[C]// British Machine Vision Conference 2008, September 2-4, 2008, Leeds, UK. [S.l.:s.n.], 2008.
- [12] JAIN P, KULIS B, GRAUMAN K. Fast image search for learned metrics[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. Piscataway: IEEE Press, 2009: 1-8.
- [13] JI J, LI J, YAN S, et al. Super-bit locality-sensitive hashing[C]// International Conference on Neural Information Processing Systems, Dec 3-6, 2012, Lake Tahoe, Nevada, USA. New York: ACM Press, 2012: 108-116.
- [14] RAGINSKY M, LAZEBNIK S. Locality-sensitive binary codes from shift-invariant kernels.[C]// Advances in Neural Information Processing Systems 2009, Dec 7-10, 2009, Vancouver, British Columbia, Canada. [S.l.:s.n.], 2009: 1509-1517.
- [15] GONG Y, KUMAR S, VERMA V, et al. Angular quantization-based binary codes for fast similarity search[J]. Advances in Neural Information Processing Systems, 2012: 1196-1204.
- [16] HE K, WEN F, SUN J. K-means hashing: an affinity-preserving quantization method for learning binary compact codes[C]// Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. Piscataway: IEEE Press, 2013: 2938-2945.
- [17] KULIS B, JAIN P, GRAUMAN K. Fast similarity search for learned metrics[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(12): 2143-57.
- [18] LIN G, SHEN C, SHI Q, et al. Fast supervised hashing with decision trees for high-dimensional data[C]// Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. Piscataway: IEEE Press, 2014: 1971-1978.
- [19] NOROUZI M, FLEET D J. Minimal loss hashing for compact binary codes[C]// International Conference on International Conference on Machine Learning, June 28-July 2, 2011, Bellevue, Washington, USA. New York: ACM Press, 2011: 353-360.
- [20] SHEN F, SHEN C, SHI Q, et al. Inductive hashing on manifolds[C]// IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. Piscataway: IEEE Press, 2013: 1562-1569.
- [21] WANG J, KUMAR S, CHANG S F. Semi-supervised hashing for large-scale search[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(12): 2393-2406.
- [22] KULIS B, GRAUMAN K. Kernelized locality-sensitive hashing for scalable image search[C]// IEEE International Conference on Computer Vision, Sept 29-Oct 2, 2009, Kyoto, Japan. Piscataway: IEEE Press, 2010: 2130-2137.
- [23] RAHIMI A, RECHT B. Random features for large-scale kernel

- machines[C]// International Conference on Neural Information Processing Systems, Dec 3-6, 2007, Vancouver, British Columbia, Canada. New York: ACM Press, 2007: 1177-1184.
- [24] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436.
- [25] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada, USA. New York: ACM Press, 2012: 1097-1105.
- [26] SALAKHUTDINOV R, HINTON G. Semantic hashing[J]. International Journal of Approximate Reasoning, 2009, 50(7): 969-978.
- [27] SALAKHUTDINOV R, HINTON G. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [28] GOLDBERGER J, ROWEIS S, HINTON G E, et al. Neighborhood components analysis[C]// International Conference on Neural Information Processing Systems, Nov 22-25, Vancouver, British Columbia, Canada. Cambridge: MIT Press, 2004: 513-520.
- [29] TORRALBA A, FERGUS R, WEISS Y. Small codes and large image databases for recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. Piscataway: IEEE Press, 2008: 1-8.
- [30] MASCI J, BRONSTEIN A M, BRONSTEIN M M, et al. Sparse similarity-preserving hashing[J]. Computer Science, 2013.
- [31] GREGOR K, LECUN Y. Learning fast approximations of sparse coding[C]// International Conference on Machine Learning, June 21-24, 2010, Haifa, Israel. New York: ACM Press, 2010: 399-406.
- [32] LIONG V E, LU J, WANG G, et al. Deep hashing for compact binary codes learning[C]// Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. Piscataway: IEEE Press, 2015: 2475-2483.
- [33] LIU W, WANG J, JI R, et al. Supervised hashing with kernels[C]// IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. Piscataway: IEEE Press, 2012: 2074-2081.
- [34] XIA R, PAN Y, LAI H, et al. Supervised hashing for image retrieval via image representation learning[J]. Association for the Advance of Artificial Intelligence, 2014: 2156-2162.
- [35] LAI H, PAN Y, LIU Y, et al. Simultaneous feature learning and hash coding with deep neural networks[J]. Computer Vision and Pattern Recognition, 2015: 3270-3278.
- [36] ZHAO F, HUANG Y, WANG L, et al. Deep semantic ranking based hashing for multi-label image retrieval[C]// Computer Vision and Pattern Recognition, Sept 11-14, 2015, Monterey, CA, USA. Piscataway: IEEE Press, 2015: 1556-1564.
- [37] ZHANG R, LIN L, ZHANG R, et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4766-4779.
- [38] LIU H, WANG R, SHAN S, et al. Deep supervised hashing for fast image retrieval[C]// Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. Piscataway: IEEE Press, 2016: 2064-2072.
- [39] LIN K, YANG H F, HSIAO J H, et al. Deep learning of binary hash codes for fast image retrieval[C]// Computer Vision and Pattern Recognition Workshops, June 7-12, 2015, Boston, MA, USA. Piscataway: IEEE Press, 2015: 27-35.
- [40] YANG H F, LIN K, CHEN C S. Supervised learning of semantics-preserving hash via deep convolutional neural networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015 (99): 1.
- [41] ZHANG J, PENG Y, YUAN M. SCH-GAN: semi-supervised cross-modal hashing by generative adversarial network[J]. arXiv: 1802.02488v1.
- [42] QIAN J, ZHU Q, CHEN H. Multi-granularity locality-sensitive bloom filter[J]. IEEE Communications Letters, 2015, 64(12): 3500-3514.

[作者简介]



袁明汶 (1993-), 男, 宁波大学信息科学与工程学院硕士生, 主要研究方向为机器学习、人工智能、大数据检索。



钱江波 (1974-), 男, 博士, 宁波大学信息科学与工程学院教授, 主要研究方向为数据处理与挖掘、机器学习、多维索引与查询优化。



董一鸿 (1969-), 男, 博士, 宁波大学信息科学与工程学院教授, 主要研究方向为大数据、数据挖掘和人工智能。

陈华辉 (1964-), 男, 博士, 宁波大学信息科学与工程学院教授, 主要研究方向为数据处理与挖掘、云计算。