软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(3):742−752 [doi: 10.13328/j.cnki.jos.006178]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

捕获局部语义结构和实例辨别的无监督哈希*

李长升[1], 闵齐星[2], 成雨蓉[1], 袁 野[1], 王国仁[1]

[1](北京理工大学 计算机学院,北京 100081)

[2](电子科技大学 计算机科学与工程学院,四川 成都 611731)

通讯作者: 李长升, E-mail: changshengli507@163.com

摘 要: 由于具有低存储成本、高效检索、低标注成本等方面的优势,无监督的哈希技术已经引起了学术界越来越多的关注,并且已经广泛地应用到大规模数据库检索问题中.先前的无监督方法大部分依靠数据集本身的语义结构作为指导信息,要求在哈希空间中,数据的语义信息能够得到保持,从而完成哈希编码的学习.因此,如何精确地表示语义结构以及哈希编码成为了无监督哈希方法成功的关键.提出一种新的基于自监督学习的策略进行无监督哈希编码学习.具体来讲,首先利用对比学习在目标数据集上对网络进行学习,从而能够构建准确的语义相似性结构;接着,提出一个新的目标损失函数,期望在哈希空间中,数据的局部语义相似性结构能够得到保持,同时,哈希编码的辨识力能够得到提升,提出的网络框架是端到端可训练的;最后,提出的算法在两个大规模图像检索数据集上进行了测试,大量的实验验证了所提出算法的有效性.

关键词: 无监督哈希;对比学习;实例辨别;局部语义结构

中图法分类号: TP311

中文引用格式: 李长升,闵齐星,成雨蓉,袁野,王国仁.捕获局部语义结构和实例辨别的无监督哈希.软件学报,2021,32(3):742−752. http://www.jos.org.cn/1000-9825/6178.htm

英文引用格式: Li CS, Min QX, Cheng YR, Yuan Y, Wang GR. Local semantic structure captured and instance discriminated by unsupervised Hashing. Ruan Jian Xue Bao/Journal of Software, 2021,32(3):742−752 (in Chinese). http://www.jos.org.cn/1000-9825/6178.htm

Local Semantic Structure Captured and Instance Discriminated by Unsupervised Hashing

LI Chang-Sheng[1], MIN Qi-Xing[2], CHENG Yu-Rong[1], YUAN Ye[1], WANG Guo-Ren[1]

[1](School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

[2](School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: Recently, unsupervised Hashing has attracted much attention in the machine learning and information retrieval communities, due to its low storage and high search efficiency. Most of existing unsupervised Hashing methods rely on the local semantic structure of the data as the guiding information, requiring to preserve such semantic structure in the Hamming space. Thus, how to precisely represent the local structure of the data and Hashing code becomes the key point to success. This study proposes a novel Hashing method based on self-supervised learning. Specifically, it is proposed to utilize the contrast learning to acquire a compact and accurate feature representation for each sample, and then a semantic structure matrix can be constructed for representing the similarity between samples. Meanwhile, a new loss function is proposed to preserve the semantic information and improve the discriminative ability in the Hamming space, by the spirit of the instance discrimination method proposed recently. The proposed framework is end-to-end trainable. Extensive

* 基金项目: 国家自然科学基金(61806044, U2001211, 61932004, 61732003); 北京理工大学青年教师学术启动计划(3070012222010)

Foundation item: National Natural Science Foundation of China (61806044, U2001211, 61932004,61732003); Beijing Institute of Technology Research Fund Program for Young Scholars (3070012222010)

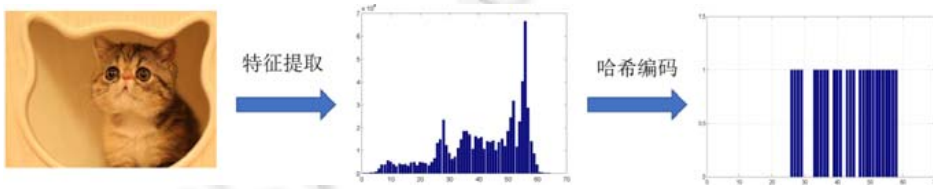
本文由“支撑人工智能的数据管理与分析技术”专刊特约编辑陈雷教授、王宏志教授、童咏昕教授、高宏教授推荐.

收稿时间: 2020-07-20; 修改时间: 2020-09-03; 采用时间: 2020-11-06; jos 在线出版时间: 2021-01-21

experiments on two large-scale image retrieval datasets show that the proposed method can significantly outperform current state-of-the-art methods.

Key words: unsupervised Hashing; contrast learning; instance discrimination; local semantic structure

随着互联网的快速发展以及数据(例如图片、视频、文档等)的爆炸式增长,如何快速地检索到用户需要的信息,已经成为学术界和工业界研究的热点问题之一.作为已经被公认为是一种非常高效地用于大规模信息检索的手段之一,哈希技术近年来得到了突飞猛进的发展.从原理上来讲,哈希方法通常将高维连续空间的数据(例如图像、视频、文本等)映射到一个低维的二进制空间中(也就是,哈希空间),如图 1 所示.在映射的过程中,期望在哈希空间中能够保持原始空间的信息.由于使用二进制编码对数据进行特征表示,哈希方法可以极大地减少存储代价以及计算复杂度,并因此可以快速地对大规模数据集进行检索查询.因此,哈希方法可以被视为一种支持大规模数据检索的高效特征学习的新技术.由于其具有广泛的潜力,目前为止,哈希方法已经被广泛地应用于各种各样的任务中,包括跨媒体检索[1]、推荐系统[2]、复制检测[3]等.



(a) 原始图像 (b) 颜色直方图 (c) 哈希代码

Fig.1 Brief introduction of Hashing

图 1 哈希过程的简单介绍

早先的哈希方法大部分是不依赖于数据的,例如,经典的局部敏感哈希(locality sensitive Hashing,简称 LSH)[4]试图通过随机映射的方式产生嵌入表示.这类技术的一个优势是在极限情况下,随着哈希编码位数的增加,随机映射可以保持输入间的距离.由于这类方法产生哈希函数不依赖于数据集本身,因此得到的哈希函数未必是全局最优的.近年来,数据依赖的哈希方法在学术界得到了更多的关注,并得到迅速发展[5].这类方法对目标数据集进行学习建模,从而产生更为精确简洁的哈希编码.由于数据依赖的方法常常获得令人满意的效果,因此各种各样的哈希方法逐渐被提出.从哈希编码学习过程中是否有监督信息介入的角度进行划分,现有的哈希方法大致可分为监督的方法[6−8]、半监督的方法[9,10]和无监督的方法[11−17].监督的哈希方法通常利用监督的信息(例如标签信息)进行哈希编码学习.监督的信息包括单个样本的标签信息、成对样本的标签信息以及序列标签信息.代表性的方法包括监督离散哈希(supervised discrete Hashing,简称 SDH)[18]、快速监督哈希(fast supervised Hashing,简称 FastH)[6]、基于排序的监督哈希(ranking-based supervised Hashing,简称 RSH)[19]、卷积神经网络哈希(convolutional neural network Hashing,简称 CNNH)[20].监督的方法存在着一些问题:首先,监督的方法通常要求哈希函数具有较强的辨识力,否则难以保证模型的性能.实际场景中的数据相对复杂,往往需要较多的编码来保证模型的精度,然而这无形之中增加了存储的空间.Wang 等人[9]提出了半监督的哈希方法,他们通过对数据对进行学习,保证相似的数据对的哈希编码仍然是相似的,不相似的数据对其哈希编码仍然是不相似的,同时要求哈希编码在没有标签的数据上的信息熵最大化.Mu 等人[10]对部分数据进行人工标注,标注数据对为语义上相似的数据对和语义上不相似的数据对,利用二次规划对问题进行求解,从而获得较为精确地哈希函数.无监督的哈希方法没有利用任何的监督信息,而是仅仅利用数据的特征信息进行学习训练.代表性的工作包括迭代量化方法(iterative quantization,简称 ITQ)[11]、离散图哈希(discrete graph Hashing,简称 DGH)[13]、大规模图哈希(scalable graph Hashing,简称 SGH)[21]等等.无监督哈希在学习过程中由于没有使用标签信息,节省了数据标注的成本,因此简单易实现.然而,也正因为没有涉及监督信息,无监督哈希问题是非常具有挑战性的.因此,本文主要研究无监督哈希问题.

在过去几年,尽管有大量的无监督哈希方法相继被学者提出来,然而这些方法仍存在着下面的问题:1) 由于数据是没有标签信息的,如何精确地构建数据间的语义相似结构仍然是一个开放的问题;2) 在哈希编码的学习过程中,大部分方法仅仅试图去保持数据的语义结构,然而忽视了哈希编码的辨别力.众所周知,数据特征表征的辨别力对于下游任务起着非常关键的作用[22],因此,如何提高哈希编码的辨别力是值得探索的.

基于以上情况并受到实例分辨力工作[23]的启发,本文提出了一种新的深度无监督的哈希学习方法:通过自监督学习对数据的语义相似性结构进行精确描述;提出一个新的目标损失函数,期望在哈希空间中,数据的语义结构能够得到保持,同时能够提升哈希编码的辨识力.另外,本文增加了一个规则项以期减少引入松弛带来的损失.本文提出的框架是端到端可训练的,并采用标准的反向传播算法进行优化.本文对图像分类模型 VGG-F 模型[24]做了以上改造及训练,通过在 FLICKR25K 和 NUSWIDE 两个常用的数据集上与目前流行的无监督哈希学习方法进行检索实验对比,证明了本方法的可行性与有效性

本文第 1 节主要介绍目前已有的哈希学习算法及其分类,同时介绍现有的无监督哈希学习存在的问题以及本文工作的主要技术路线.第 2 节从问题定义、网络架构、损失函数等几个方面详细阐述本文提出的模型性能提升方法.第 3 节在两个常用数据集上证明本方法能够在不同哈希编码长度的条件下,均能够提高模型的检索精度.第 4 节对本文工作做出总结,并给出未来的工作展望.

1 相关工作

面向大规模数据集设计高效的特征学习算法,对于检索具有十分重要的应用价值.在构建高效的大规模检索系统时,往往存在着两个最主要的问题:数据的存储成本和检索速度.当前,文本、图像、视频等多媒体数据往往具有高维度的特征,因此检索方法面临着“特征维数灾难”的严峻挑战,使得系统的存储空间、计算复杂度都急剧增加,从而影响了检索系统的性能.为了解决上面的难题和挑战,哈希学习技术被提出来,并成为信息检索领域和机器学习领域的研究热点.哈希学习(learning to Hash)[5]对数据自身的特点和结构进行分析,依靠机器学习的方法将高维连续数据映射为哈希编码(也就是二进制串的形式),同时在哈希空间中尽可能地保持原空间中的结构信息.由于其二进制表示形式,哈希学习能够显著减少数据的存储成本以及计算复杂度,从而有效提高检索系统的效率.由于本文主要研究无监督哈希方法,因此本节主要对无监督哈希学习方法进行回顾总结.

传统的无监督哈希学习方法通常基于浅层的结构进行哈希编码学习,这些方法通常将特征学习和哈希编码当作两个分开的过程.代表性的算法包括 ITQ[11]等.ITQ 试图先对原始空间的数据集用 PCA 进行降维处理,将数据集中的数据点映射到一个二进制超立方体的顶点上,使得对应的量化误差最小,从而得到对应该数据集较为精确的哈希编码.近几年,随着深度学习在各种视觉任务和机器学习中取得了令人惊讶的效果,深度学习也逐渐被应用到哈希学习中,例如语义哈希(semantic Hashing)[25]、深度自编码哈希(deep auto-encoder Hashing)[26]和深度二进制描述子(deep binary descriptors,简称 DeepBit)[27].语义哈希使用预训练的限制玻尔兹曼机构建自编码网络,从而能够产生有效的哈希编码,并且能够准确地重构原始输入.深度自编码哈希设计了一个非常深的自编码器用于映射原始输入到哈希空间中,并且利用重构损失指导哈希编码的学习.深度二进制描述子将特征学习和哈希编码学习融合到一个框架中,并取得了不错的效果.

2 提出的方法

本文提出了一种基于局部语义结构和实例辨别的深度无监督哈希方法.本文认为:在哈希编码学习过程中,提升哈希编码的分辨力可以提高模型的表达能力和检索能力.该方法主要包含两个部分:一是利用对比学习(contrastive learning)对局部语义相似结构进行提炼,使其具有不仅能够表示数据的语义信息,同时能够表示数据的辨识信息;二是提出一个新的目标损失函数,在哈希空间中利用对比学习,使得哈希编码不仅能够保持数据的语义信息,同时提升哈希编码的辨识能力.下面将分别从问题定义、网络架构、语义结构矩阵以及哈希编码学习等几个方面进行具体的阐述.

2.1 问题定义

首先,本文给出一些主要符号的表示,见表 1.

Table 1 Summarization of notations

表 1 符号表示

符号名称	符号表示含义描述
$\boldsymbol{A}$	表示矩阵
$\boldsymbol{a}_i$	表示矩阵的第 i 列
$\boldsymbol{a}_{ij}$	表示矩阵的第 i 行第 j 列元素
$\|\|A\|\|_F$	表示矩阵的 Frobenius 范数
A^T	表示矩阵 $\boldsymbol{A}$ 的转置
$\exp(\cdot)$	表示指数操作
$\boldsymbol{a}\cdot\boldsymbol{b}$	表示两个向量的内积
$\|\boldsymbol{A}\|$	表示对矩阵每个元素求绝对值
L	表示哈希编码长度
N	表示样本的个数
d	表示样本的原始特征维度

给定一组训练数据 $X=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n]\in\mathbb{R}^{d\times n}$,本文的目标是学习一组二进制哈希编码:

$$B=[\boldsymbol{b}_1,\boldsymbol{b}_2,\ldots,\boldsymbol{b}_n]\in\{-1,1\}^{L\times n}.$$

为了达到这个目的,本文试图求解一组有效的哈希函数,如下式所示:

$$\boldsymbol{b}_i=[h_1(\boldsymbol{x}_i),\ldots,h_L(\boldsymbol{x}_i)]=[\mathrm{sgn}(F(\boldsymbol{x}_i;\boldsymbol{W}_1)),\ldots,\mathrm{sgn}(F(\boldsymbol{x}_i;\boldsymbol{W}_L))] \tag{1}$$

其中,$\boldsymbol{W}_1,\ldots,\boldsymbol{W}_L$ 表示模型学习的参数.$\mathrm{sgn}(\cdot)$表示符号函数,定义为

$$\mathrm{sgn}(x)=\begin{cases}1, & \text{if } x\geqslant 0\\ -1, & \text{if } x<0\end{cases} \tag{2}$$

受深度学习在哈希学习方法中突出的表现激励[27,28],为了能够较好地将原始数据映射到哈希空间,本文仍然采用了深度神经网络作为基本架构对哈希函数进行学习.尽管先前许多方法试图在哈希空间中保持数据的结构信息,然而他们忽视了哈希编码的辨识力.因此,本文目的是在哈希空间中不仅保持数据的语义相似结构,而且试图提升哈希编码的辨识力.

2.2 模型学习

为了完成上面的目标,本文提出了基于实例辨识力的框架用于哈希编码学习,如图 2 所示.

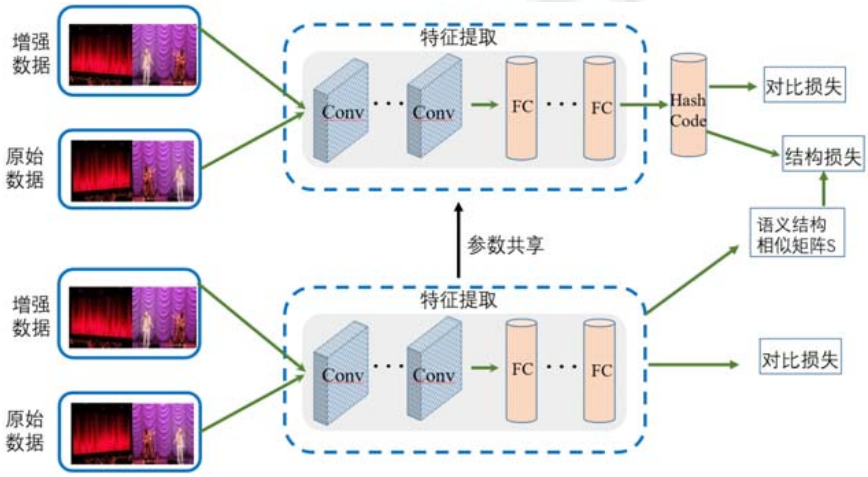


Fig.2 The architexture of the proposed method

图 2 本文提出方法的框架

整个框架主要分成两个部分:(1) 构建语义结构相似矩阵;(2) 哈希编码学习.具体地,为了构建相似矩阵 $\boldsymbol{S}$,本文首先利用对比学习的策略对目标数据集进行训练,使得学习到的特征具有一定的辨识力.模型更新结束后,利用网络中间层的特征作为数据新的特征表示.基于新的特征表示构建语义相似结构 $\boldsymbol{S}$;为了学习哈希编码,首先利用结构损失函数优化网络,试图在哈希空间中保持数据的语义局部结构;同时,对原始数据和增强数据组成训练样本对,使用对比损失增强特征表达的辨识力.

为了构建语义局部结构矩阵 $\boldsymbol{S}$,本文利用 VGG-F[24]模型作为卷积主干结构进行特征提取.由于本文研究的是无监督哈希学习问题,因此数据的标签信息是不可得的.为了能够训练网络,本文采用自监督学习的机制构建辅助任务对网络进行学习.本文使用如下的损失函数:

$$L_c = -\log \frac{\exp(\boldsymbol{x}_i^* \cdot \boldsymbol{x}_i^+ / \tau)}{\sum_{j=1}^{m} \exp(\boldsymbol{x}_i^* \cdot \boldsymbol{x}_j^+ / \tau)} \tag{3}$$

其中,τ是一个超参数. $\boldsymbol{x}_i^*$ 和 $\boldsymbol{x}_i^+$ 是 $\boldsymbol{x}_i$ 的两个增强样本,例如通过随机旋转、加噪音等方式对原图像进行数据增强.等式(3)的目的是以数据的两个增强样本构成正样本对,同时以这个数据的增强样本与其他数据的增强样本构成负样本对,以此训练一个分类器,试图将同一样本的增强样本分类到同一类中去.通过上述辅助任务,网络学习到的特征能够具有一定的辨识力.注意:为了防止数据过拟合,本文没有使用原始数据去更新网络,仅仅使用了原始数据的增强样本去更新网络.

当网络更新停止后,本文利用 fc-7 层的特征作为数据新的特征表示,并构建如下的结构矩阵:

$$\boldsymbol{S}_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{x}_j \in \Theta_k(\boldsymbol{x}_i) \\ 0, & \text{if } \boldsymbol{x}_j \notin \Theta_k(\boldsymbol{x}_i) \cup \Omega_k(\boldsymbol{x}_i) \\ -1, & \text{if } \boldsymbol{x}_j \in \Omega_k(\boldsymbol{x}_i) \end{cases} \tag{4}$$

其中,$\Theta_k(\boldsymbol{x}_i)$表示数据点 $\boldsymbol{x}_i$ 的 K 个最近邻点,$\Omega_k(\boldsymbol{x}_i)$表示所有数据点中离着数据点 $\boldsymbol{x}_i$ 最远的 K 个数据点.在等式(4)中,如果两个点是近邻点,那么认为他们的语义信息是相似的,因此,这两个数据点在哈希空间中的距离应该是比较小的;如果两个点的距离较远,那么认为他们的语义信息是不相似的,因此,这两个数据点在哈希空间中的距离应该是比较远的.算法 1 给出了构建结构矩阵的具体步骤.

算法 1. 语义相似结构矩阵构建.

输入:训练数据 $\boldsymbol{X}$,迷你批(mini-batch)大小 m;

输出:语义相似结构矩阵 $\boldsymbol{S}$.

1: 初始化:使用在 imagenet 数据集预训练的模型初始化 VGG-F 网络的参数;
2: 重复直到收敛:
3: 从 $\boldsymbol{X}$ 中随机选择 m 个样本构建一个迷你批;
4: 对迷你批的每个数据做数据增强;
5: 通过前向传播计算损失函数(3);
6: 利用反向传播算法更新网络参数;
7: 基于更新后的 VGG-F 网络,提取 fc-7 层的特征作为数据集新的特征表示;
8: 通过等式(4),构建语义结构相似矩阵 $\boldsymbol{S}$

为了在哈希空间中保持数据的语义结构,同时提升哈希特征的辨识力,本文提出下面的目标函数:

$$\min_{\boldsymbol{B}} L_b = -\log \frac{\exp(\boldsymbol{b}_i \cdot \boldsymbol{b}_i^+ / \tau)}{\sum_{j=1}^{m} \exp(\boldsymbol{b}_i \cdot \boldsymbol{b}_j^+ / \tau)} + \alpha \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |\boldsymbol{S}_{ij}| (\boldsymbol{M}_{ij} - \boldsymbol{S}_{ij})^2 \tag{5}$$

其中,$\boldsymbol{b}_i$ 和 $\boldsymbol{b}_i^+$ 分别表示数据样本 $\boldsymbol{x}_i$ 和它的增强样本 $\boldsymbol{x}_i^+$ 的哈希编码,$\boldsymbol{b}_i=\text{sgn}(F(\boldsymbol{x}_i;\Phi))$,$\Phi$表示网络学习的参数.矩阵 $\boldsymbol{M}$ 定义为

$$\boldsymbol{M}_{ij}=\frac{1}{L}\boldsymbol{b}_i^T\boldsymbol{b}_j,\boldsymbol{b}_i\in\{-1,1\}^L \tag{6}$$

在等式(5)中,第 1 项的目的是希望数据样本与它的增强样本在哈希空间中尽可能的接近,从而使得哈希编码具有一定的辨识力;第 2 项的目的是希望数据在哈希空间中和连续特征空间中的语义结构保持一致.通过联合优化这两项,数据的语义结构能够得到保持,同时数据的辨识力得到提升.

在等式(5)中,二进制表示使得网络的优化变得十分困难.为了有效地对网络进行梯度更新,本文使用 tanh(·)函数代替 sgn(·)函数,从而对目标函数进行松弛,因此提出下面的目标函数:

$$\left.\begin{aligned}&\min_{\boldsymbol{B}}L_b=-\log\frac{\exp(\boldsymbol{b}_i\cdot\boldsymbol{b}_i^+/\tau)}{\sum_{j=1}^{m}\exp(\boldsymbol{b}_i\cdot\boldsymbol{b}_j^+/\tau)}+\alpha\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\boldsymbol{S}_{ij}|(\boldsymbol{M}_{ij}-\boldsymbol{S}_{ij})^2\\&\text{s.t. }\boldsymbol{b}_i=\tanh(F(\boldsymbol{x}_i;\boldsymbol{\Phi})),\boldsymbol{M}_{ij}=\frac{1}{L}\boldsymbol{b}_i^T\boldsymbol{b}_j\end{aligned}\right\} \tag{7}$$

另外,为了尽可能地减少上述松弛带来的损失,本文增加了另外一个规则项,使得哈希编码的值尽可能接近 1 或者−1,因此得到下面的目标函数:

$$\left.\begin{aligned}&\min_{\boldsymbol{B}}L_b=-\log\frac{\exp(\boldsymbol{b}_i\cdot\boldsymbol{b}_i^+/\tau)}{\sum_{j=1}^{m}\exp(\boldsymbol{b}_i\cdot\boldsymbol{b}_j^+/\tau)}+\alpha\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\boldsymbol{S}_{ij}|(\boldsymbol{M}_{ij}-\boldsymbol{S}_{ij})^2+\beta\sum_{i=1}^{L}(1-|\boldsymbol{b}_{ij}|)\\&\text{s.t. }\boldsymbol{b}_i=\tanh(F(\boldsymbol{x}_i;\boldsymbol{\Phi})),\boldsymbol{M}_{ij}=\frac{1}{L}\boldsymbol{b}_i^T\boldsymbol{b}_j\end{aligned}\right\} \tag{8}$$

其中,$\alpha\geqslant 0$ 和 $\beta\geqslant 0$ 是两个超参数.

为了求解等式(8),本文采用标准反向传播算法对梯度进行更新,整个训练过程见算法 2.

算法 2. 哈希编码训练.

输入:训练数据 $\boldsymbol{X}$,迷你批(mini-batch)大小 m,超参数 α 和 β;

输出:神经网络参数 $\boldsymbol{\Phi}=\{\boldsymbol{W}_1,\ldots,\boldsymbol{W}_L\}$ 和训练数据的哈希编码.

1: 初始化:使用在训练数据集上微调后的模型初始化 VGG-F 网络的参数;

2: 重复直到收敛:

3: 从 $\boldsymbol{X}$ 中随机选择 m 个样本构建一个迷你批;

4: 对迷你批的每个数据做数据增强;

5: 通过前向传播计算损失函数(8);

6: 利用反向传播算法更新网络参数;

7: 基于更新后的 VGG-F 网络,对于数据集提取网络最后一层作为数据的哈希编码;

当网络训练完成后,对于任意其他不在训练集中的数据点 $\boldsymbol{x}_t$,可以利用下式直接计算其哈希编码:

$$\boldsymbol{b}_t=\text{sgn}(F(\boldsymbol{x}_t;\boldsymbol{\Phi})) \tag{9}$$

任意数据点的哈希编码映射过程如算法 3 所示.

算法 3. 哈希编码测试.

输入:查询数据 $\boldsymbol{x}_t$,神经网络的参数 $\boldsymbol{\Phi}$;

输出:查询数据 $\boldsymbol{x}_t$ 的哈希编码 $\boldsymbol{b}_t$.

1: 利用前向传播算法计算网络的输出;

2: 使用等式(9)计算 $\boldsymbol{x}_t$ 的哈希编码;

3 实验及分析

本节验证了所提方法的有效性,包括实验平均精度均值(MAP)、参数敏感分析、消融实验.本文利用 Pytorch 实现深度哈希模型,通过动量式的批量随机梯度下降法优化模型参数,其中批量大小为 16,动量参数设置为 0.9,

学习率固定为 0.001.为了与其他哈希模型进行公平比较,本方法将原始图像直接裁剪为 224×224 的尺寸作为模型的输入,不做任何数据增强,并在 VGG-F 的 fc-7 提取特征向量.随后将特征向量输入哈希层,得到每个原始图像的哈希编码,其中,哈希层是模型最后的全连接层.

验证实验在 2 个基准数据集——NUSWIDE,FLICKR25K 上进行.在每个数据集上,都与一些效果好的深度学习方法和传统的浅层方法进行了对比分析.使用常见的评价准则来衡量本方法的效果:平均精度均值.

平均精度均值为每个询问数据(query)的精度均值(AP)的平均:

$$AP=\frac{1}{N}\sum_{r=1}^{R}P(r)\delta(r),$$

其中,N 表示为与询问数据标签相关的样本数量,$P(r)$为前 r 个检索样本的准确率,$\delta(r)$表示第 r 个检索样本是否与询问数据相关.本文设置 R 为 5 000,并且确定:若两个样本至少有一个标签相同,则这两个样本相关.

3.1 数据集及实验环境

本文实验在一个服务器节点上运行,该服务器的操作系统为 Linux version 4.4.0-116-generic (buildd@lgw01-amd64-021) (gcc version 5.4.0 20160609 (Ubuntu 5.4.0-6ubuntu1~16.04.9)),处理器为 Intel(R) Xeon(R) Silver 4210CPU@2.20GHz,内存 64GB.

FLICKR25K 数据集包含从 Flickr 网站收集到的 25 000 个图像,共分为 24 类.随机选择 2 000 个图像作为测试集,剩余图像作为检索集,并从检索集中随机选择 10 000 个图像作为训练集.

NUSWIDE 数据集包含 269 648 个图像,共有 81 种类别.本文使用的数据子集包含 10 个最常见的标签,随机选择 5 000 个图像作为测试集,剩余图像作为检索集,并从检索集中随机选择 5 000 个图像作为训练集.

3.2 FLICKR25K数据集上的实验结果

本方法与其他对比方法在 FLICKR25K 数据集上的实验结果见表 2.表 2 显示了各算法在哈希编码长度从 16 位变化到 128 位时获得的平均精度均值.可以看出:本文提出的方法在不同哈希编码长度的实验中比其他对比方法表现更好,在 16 位、32 位、64 位、128 位哈希编码长度上,本方法比表现第二好的深度无监督哈希方法,SSDH 的 MAP 分别高出 5.17%,6.46%,6.70%,7.45%,从而证明了本文方法的优势.对比方法中,ITQ[11],Spectral Hashing (SH)[29],Density Sensitive Hashing (DSH)[30],Spherical Hashing (SpH)[31],SGH[21]是传统的浅层方法,而 DeepBit[27]和 SSDH[14]是基于深度模型的方法.通过对比发现,一些非深度哈希的方法比深度哈希方法 DeepBit 的 MAP 更高.这可能是因为深度哈希方法在缺乏监督信息时,不能完全利用深度网络的特征表达能力并且容易过拟合到局部最小点,从而影响效果.本方法使用了基于对比学习的自监督哈希编码的方法,并且对哈希编码进行正则化,从而完成了最好的 MAP 结果.

Table 2 MAP of different code length in FLICKR25K

表 2 FLICKR25K 数据集上对不同哈希编码长度的测试平均精度均值

算法	FLICKR25K			
	16bits	32bits	64bits	128bits
ITQ	0.649 2	0.651 8	0.654 6	0.657 7
SH	0.609 1	0.610 5	0.603 3	0.601 4
DSH	0.645 2	0.654 7	0.655 1	0.655 7
SpH	0.611 9	0.631 5	0.638 1	0.645 1
SGH	0.636 2	0.628 3	0.625 3	0.620 6
DeepBit	0.593 4	0.593 3	0.619 9	0.634 9
SSDH	0.724 0	0.727 6	0.737 7	0.734 3
Ours	**0.775 7**	**0.792 2**	**0.804 7**	**0.808 8**

3.3 NUSWIDE数据集上的实验结果

本方法与其他对比方法在 NUSWIDE 数据集上的实验结果见表 3.表 3 显示了各算法在哈希编码长度从 16 位变化到 128 位获得的平均精度均值.

Table 3 MAP of different code length in NUSWIDE

表 3 NUSWIDE 数据集上对不同哈希编码长度的测试平均精度均值

算法	NUSWIDE			
	16bits	32bits	64bits	128bits
ITQ	0.527 0	0.524 1	0.533 4	0.539 8
SH	0.435 0	0.412 9	0.406 2	0.410 0
DSH	0.512 3	0.511 8	0.511 0	0.526 7
SpH	0.445 8	0.453 7	0.492 6	0.500 0
SGH	0.499 4	0.486 9	0.485 1	0.494 5
DeepBit	0.384 4	0.434 1	0.446 1	0.491 7
SSDH	0.637 4	0.676 8	0.682 9	0.683 1
Ours	**0.707 0**	**0.739 7**	**0.761 3**	**0.786 8**

由表 3 所示的结果可以看出:本文提出的方法在不同哈希编码长度的实验中,比其他对比方法表现更好.在 16 位、32 位、64 位、128 位哈希编码长度上,本方法比 SSDH 的 MAP 分别高出 6.96%,6.29%,7.84%,10.37%,仍然证明了本文方法的优势.哈希编码长度越长,能够编码的信息越多,因此 MAP 更高.此外,NUSWIDE 检索集大小是 FLICKR25K 检索集大小的 10 倍左右,因此检索查找的难度急剧增加,因此,MAP 值较 FLICKR25K 在相同哈希编码长度时小.

3.4 消融实验分析

本小节进行消融实验比较,从而验证提出算法每部分的有效性.首先,将本方法划分出 3 个实验元素,见表 4,分别为:是否加入局部语义结构信息、是否加入对比学习损失、是否加入正则项损失.组合不同的实验元素,得到消融实验结果,以此观察每个实验元素对结果的影响.

Table 4 Three main componets for ablation studies

表 4 消融实验中的 3 个元素

符号	含义
C1	加入局部语义结构信息
C2	加入对比学习损失
C3	加入正则项损失

在 FLICKR25K 数据集、哈希编码长度为 16 位的条件下,进行消融实验.消融实验结果见表 5.在 FLICKR25K 数据集、哈希编码长度为 32 位的条件下,进行消融实验.消融实验结果见表 6.

Table 5 MAP of our method's variants on FLICKR25K, at code length 16bits

表 5 加入不同元素的 FLICKR25K 数据集上,16 位哈希编码的 MAP 对比

方法	C1	C2	C3	MAP
Our method	√	–	–	0.743 5
	√	√	–	0.765 9
	√	√	√	0.775 7

Table 6 MAP of our method's variants on FLICKR25K, at code length 32bits

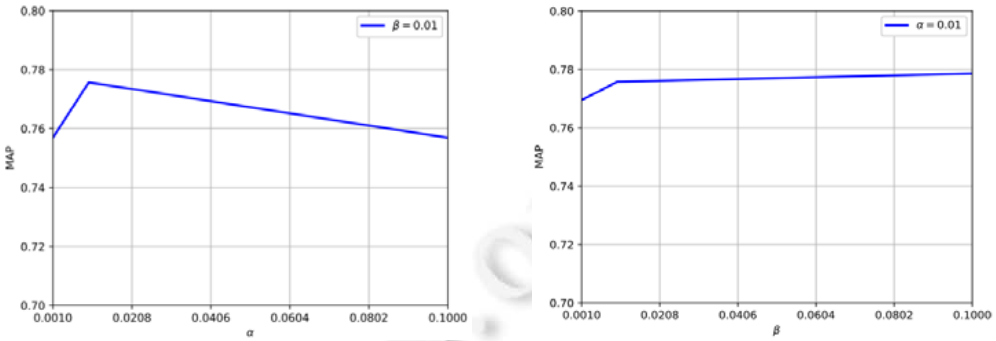
表 6 加入不同元素的 FLICKR25K 数据集上,32 位哈希编码的 MAP 对比

方法	C1	C2	C3	MAP
Our method	√	–	–	0.751 7
	√	√	–	0.789 0
	√	√	√	0.792 2

从表 5 和表 6 可以看出:加入正则项和对比学习损失项后,模型的精度均能得到提升.模型在加入语义结构信息的基础上,加入对比学习损失,通过对哈希码使用动量对比学习算法,进一步学习到更准确的哈希编码表达,极大地提高了模型的精度.在加入语义结构信息和对比学习损失的基础上,加入正则项损失,通过约束哈希码尽量趋近于 1 或–1,提升了哈希码的辨识度.因此,在哈希空间中试图保持数据的语义结构信息和提升哈希编码的辨识力,对于提升模型性能起到了积极的作用,从而验证了本文所提方法的有效性.

3.5 参数敏感性分析

本小节对所提方法中的超参数进行了敏感性分析.本文主要包含了 3 个超参数α,β,τ.本文在 FLICKR25K 数据集上哈希编码长度为 16 位的情况下进行了实验.本文首先固定正则项损失参数β为 0.01,在 0.001~0.1 内变化α,结果如图 3(a)所示.固定对比学习损失参数α为 0.01,在 0.001~0.1 内变化β,结果如图 3(b)所示.



(a) MAP 随参数α的变化 (b) MAP 随参数β的变化

Fig.3 Loss hyper-parameters of code length 16bits on FLICKR25K dataset

图 3 损失项参数对 FLICKR25K 数据集 16 位哈希编码长度实验的影响

从图 3 中可以看出:固定参数β,随着α的增加,MAP 先增加后减少,并且在α为 0.01 时表现最好;固定参数α为 0.01,随着β的增加,MAP 先增加后减少,并且在β为 0.1 时表现最好.对比α和β对实验结果 MAP 的影响可以看到:对比学习损失的参数α对实验结果影响更大,而正则项损失的参数β能在一定程度上提高实验结果.本文在其他实验中固定α=0.01 和β=0.01.

固定损失项参数α=0.01 和β=0.01,在 0~0.5 内变化温度参数τ,结果如图 4 所示.

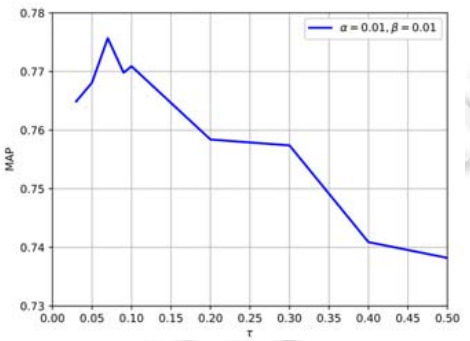


Fig.4 Temporature hyper-parameter of code length 16bits on FLICKR25K dataset

图 4 温度参数对 FLICKR25K 数据集 16 位哈希编码长度实验的影响

τ是控制数据分布集中程度的温度参数.从图 4 中可以看出:固定参数α和β,随着τ的增加,MAP 整体趋势先增加后减少,并且在τ为 0.07 时表现最好.本文的实验中,固定α=0.01,β=0.01 以及τ=0.07.

3.6 时间复杂度分析

时间复杂度即模型的运算次数,可用 FLOP(floating-point operation)衡量,表示浮点运算次数.

所有卷积层的时间复杂度为

$$O\left(\sum_{l=1}^{d} n_{l-1}\cdot s_l^2\cdot n_l\cdot m_l^2\right),$$

其中,l 表示卷积层的下标,d 表示卷积层的数量,n_l 表示第 l 层网络卷积核的数量,n_{l-1} 表示第 l 层网络的输入通道数,s_l 表示卷积核的边长,m_l 表示输出特征图的边长.训练过程和测试过程的时间复杂度不同,每个图像的训练用时大约是测试用时的 3 倍(前向传播一倍,反向传播两倍).本文计算了哈希编码训练过程中的 VGG-F 网络与哈希编码层的时间复杂度共为 31.0GFlops,其中,全连接层与池化层的时间开销占总的时间复杂度的 0.8%.

4 结论及展望

针对无监督的哈希学习问题,本文提出了一种基于语义结构保持和实例分辨力的深度无监督哈希学习的框架.为了能够提升哈希编码的辨识力,采用自监督学习策略,不仅对语义结构进行学习,同时也指导哈希编码的学习.本文在两个常用于评估哈希方法的数据集上进行了实验,广泛的实验设计与分析验证了提出方法的有效性.下一步研究工作的重点将是尝试设计更加有效的自监督学习任务以及更加有效的自监督学习损失函数;同时,如何优化算法提升模型的训练速度,也是下一步工作的重点.

References:

[1] Wu GS, Lin ZJ, Ha JG, Liu L, Ding GG, Zhang BC, Shen JL. Unsupervised deep Hashing via binary latent factor models for large-scale cross-modal retrieval. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence (IJCAI 2018). 2018. 2854−2860.

[2] Aytekin AM, Aytekin T. Real-time recommendation with locality sensitive Hashing. Journal of Intelligent Information Systems, 2019,53(1):1−26.

[3] Sun JD, Wang J, Yuan H, Liu XC, Liu J. Unequally weighted video Hashing for copy detection. In: Li SP, ed. Proc. of the 19th Int'l Conf. on Multimedia Modeling, Advances in Multimedia Modeling (MMM 2013). Part I. Huangshan, Berlin, Heidelberg: Springer-Verlag, 2013. 546−557.

[4] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via Hashing. In: Atkinson MP, ed. Proc. of the 25th Int'l Conf. on Very Large Data Bases (VLDB'99). Edinburgh: Morgan Kaufmann publishers, 1999,99(6):518−529.

[5] Wang JD, Zhang T, Song JK, Sebe N, Shen HT. A survey on learning to Hash. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2017,40(4):769−790.

[6] Lin GS, Shen CH, Shi QF, van den Hengel A, Suter D. Fast supervised Hashing with decision trees for high-dimensional data. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2014). Columbus: IEEE Computer Society, 2014. 1963−1970.

[7] Liu W, Wang J, Ji RR, Jiang YG, Chang SF. Supervised Hashing with kernels. In: Proc. of the 2012 IEEE Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2012. 2074−2081.

[8] Shen FM, Zhou X, Yang Y, Song JK, Shen HT, Tao DC. A fast optimization method for general binary code learning. IEEE Trans. on Image Processing, 2016,25(12):5610−5621.

[9] Wang J, Kumar S, Chang SF. Semi-Supervised Hashing for large-scale search. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(12):2393−2406.

[10] Mu Y, Shen J, Yan S. Weakly-supervised Hashing in kernel space. In: Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2010). San Francisco: IEEE Computer Society, 2010. 3344−3351.

[11] Gong YC, Lazebnik S, Gordo A, Perronnin F. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,35(12):2916−2929.

[12] Hu MQ, Yang Y, Shen FM, Xie N, Shen HT. Hashing with angular reconstructive embeddings. IEEE Trans. on Image Processing, 2017,27(2):545−555.

[13] Liu W, Mu C, Kumar S, Chang SF. Discrete graph Hashing. In: Ghahramani Z, ed. Proc. of the Advances in Neural Information Processing Systems 27: Annual Conf. on Neural Information Processing Systems 2014. 2014. 3419−3427.

[14] Yang EK, Deng C, Liu TL, Liu W, Tao DC. Semantic structure-based unsupervised deep Hashing. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence (IJCAI 2018). 2018. 1064−1070.

[15] Kumar S, Udupa R. Learning Hash functions for cross-view similarity search. In: Walsh T, ed. Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence (IJCAI 2011). Barcelona: IJCAI/AAAI, 2011. 1360−1365.

[16] Song JK, Yang Y, Yang Y, Huang Z, Shen HT. Inter-Media Hashing for large-scale retrieval from heterogeneous data sources. In: Ross KA, ed. Proc. of the ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD 2013). New York: ACM, 2013. 785−796.

[17] Zhu XF, Huang Z, Shen HT, Zhao X. Linear cross-modal Hashing for efficient multimedia search. In: Jaimes A, ed. Proc. of the ACM Multimedia Conf. (MM 2013). Barcelona: ACM, 2013. 143−152.

[18] Shen FM, Shen CH, Liu W, Shen HT. Supervised discrete Hashing. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2015). Boston: IEEE Computer Society, 2015. 37−45.

[19] Wang J, Liu W, Sun AX, Jiang YG. Learning Hash codes with listwise supervision. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV 2013). Sydney: IEEE Computer Society, 2013. 3032−3039.

[20] Xia RK, Pan Y, Lai HJ, Liu C, Yan SC. Supervised Hashing for image retrieval via image representation learning. In: Brodley CE, ed. Proc. of the 28th AAAI Conf. on Artificial Intelligence. Quebec City: AAAI, 2014. 2156−2162.

[21] Jiang QY, Li WJ. Scalable graph Hashing with feature transformation. In: Yang Q, ed. Proc. of the 24th Int'l Joint Conf. on Artificial Intelligence (IJCAI 2015). Buenos Aires: AAAI, 2015. 2248−2254.

[22] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(8):1798−1828.

[23] He KM, Fan HQ, Wu YX, Xie SN, Girshick RB. Momentum contrast for unsupervised visual representation learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR 2020). Seattle: IEEE, 2020. 9729−9738.

[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[25] Salakhutdinov R, Hinton G. Semantic Hashing. Int'l Journal of Approximate Reasoning, 2009,50(7):969−978.

[26] Krizhevsky A, Hinton GE. Using very deep autoencoders for content-based image retrieval. In: Proc. of the 19th European Symp. on Artificial Neural Networks (ESANN 2011). 2011.

[27] Lin K, Lu JW, Chen CS, Zhou J. Learning compact binary descriptors with unsupervised deep neural networks. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas: IEEE Computer Society, 2016. 1183−1192.

[28] Li C, Deng C, Li N, Liu W, Gao XB, Tao DC. Self-Supervised adversarial Hashing networks for cross-modal retrieval. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2018). Salt Lake City: IEEE Computer Society, 2018. 4242−4251.

[29] Weiss Y, Torralba A, Fergus R. Spectral Hashing. In: Koller D, ed. Proc. of the Advances in Neural Information Processing Systems 21, 22nd Annual Conf. on Neural Information Processing Systems. Vancouver: Curran Associates, Inc., 2008. 1753−1760.

[30] Jin ZM, Li C, Lin Y, Cai D. Density sensitive Hashing. IEEE Trans. on Cybernetics, 2013,44(8):1362−1371.

[31] Heo JP, Lee YW, He JF, Chang SF, Yoon SE. Spherical Hashing. In: Proc. of the 2012 IEEE Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2012.

李长升(1985－),男,博士,教授,博士生导师,主要研究领域为机器学习.

闵齐星(1996－),女,学士,主要研究领域为视频行为分析.

成雨蓉(1989－),女,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为数据库.

袁野(1981－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库.

王国仁(1966－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为数据库.