

一种基于子空间学习的图像语义哈希索引方法*

毛晓蛟, 杨育彬

(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210093)

通讯作者: 杨育彬, E-mail: yangyubin@nju.edu.cn, http://www.nju.edu.cn

摘要: 随着数据量的不断增加,快速而准确的索引算法对信息检索而言变得十分重要.针对上述问题,提出了一种基于子空间学习的索引算法.首先,利用部分有标签的数据进行子空间学习,在学习过程中,为了保证语义相同的样本在索引后保持局部性,以样本近邻间的距离衡量类内聚合度;同时,为了保证不同语义的样本在索引后增强判别性,以不同语义样本中心之间的距离衡量类间离散度.通过放松限制,用类似线性判别分析的方法进行子空间学习,将子空间作为哈希函数的投影向量.利用学习到的投影向量进一步计算偏移量,得到哈希函数.分别在数据集 MNIST 和 CIFAR-10 上进行编码判别性实验和局部性保留实验,并与相关方法进行比较,得到了较好的效果.实验结果表明该方法是有效的.

关键词: 哈希函数;子空间;偏移量;局部性保留;判别性

中图法分类号: TP391

中文引用格式: 毛晓蛟,杨育彬.一种基于子空间学习的图像语义哈希索引方法.软件学报,2014,25(8):1781-1793. <http://www.jos.org.cn/1000-9825/4488.htm>

英文引用格式: Mao XJ, Yang YB. Semantic hashing with image subspace learning. Ruan Jian Xue Bao/Journal of Software, 2014, 25(8): 1781-1793 (in Chinese). <http://www.jos.org.cn/1000-9825/4488.htm>

Semantic Hashing with Image Subspace Learning

MAO Xiao-Jiao, YANG Yu-Bin

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210093, China)

Corresponding author: YANG Yu-Bin, E-mail: yangyubin@nju.edu.cn, <http://www.nju.edu.cn>

Abstract: With the increasing amount of data being collected, developing fast indexing methods with high accuracy becomes important for information retrieval tasks. To address this issue, this paper proposes an indexing method based on hashing mechanism with subspace learning. Firstly, the subspace is learned on a set of labeled data. To guarantee the locality preserving characteristics in the original space for the samples with similar semantic labels, the distances between the nearest neighbors are computed to measure the intra-class scatter. Besides, the distances between the centers of samples with dissimilar semantic labels are also computed to measure the inter-class scatter in order to enhance the discriminative power of the codes. The projections of the hash functions are then learned by relaxing the constraint of the formula. The biases are further learned based on the projections. Finally, the proposed method is evaluated on the datasets MNIST and CIFAR-10 to compare with the state-of-the-art methods. Experimental results show that the proposed method achieves significant performance and high effectiveness in searching semantically similar neighbors.

Key words: hash function; subspace; bias; locality preserving; discriminant

随着互联网的快速发展,各种网络资源日益丰富,网络数据规模也以指数级的速度增长.在互联网上存在的各种类型数据中,图像、视频和文本占据了大部分.无论是其中哪种类型的数据,目前都达到了海量规模.2008年,

* 基金项目: 国家自然科学基金(61273257, 61321491, 61035003); 国家重点基础研究发展计划(973)(2010CB327903); 教育部新世纪优秀人才计划(NCET-11-0213); 江苏省六大人才高峰计划(2013-XXRJ-018); 江苏省自然科学基金(BK2011005)

收稿时间: 2013-03-28; 修改时间: 2013-07-12, 2013-08-02; 定稿时间: 2013-08-27

谷歌统计到的可索引网页数量已超过了1万亿^[1],2010年,著名网站Flickr统计的图片总量超过了50亿^[2].这些数据还在持续以惊人的速度增长,若干年后将达到难以想象的规模.毫无疑问,从如此大的数据库中快速、准确地搜索用户需要的数据变得十分重要,同时也存在巨大的困难.例如,给定一幅图像,如何在大规模的数据库中快速并且准确地搜索到与给定图像相似的图像.一个基本的思路是对数据库中所有图像提取特征,然后定义特征相似性度量,对数据库进行线性搜索并根据相似度排序,从而得到检索结果.然而,这样的思路存在两个重要的问题:1) 图像特征往往是一种高维化的数据,高维数据的存储要求高,计算效率和类间区分度较低;2) 对大规模数据进行线性搜索很难满足效率要求.因此,如何对高维数据进行有效索引成为一个亟需解决的问题.这方面的研究近年来得到了广泛关注,研究人员也提出了一些针对图像高维特征进行索引的算法,主要以树形结构索引算法和语义哈希索引算法为代表.使用最广泛的传统树结构索引是以KD树^[3]为基础的维度分割方法,但当向量维度增加时,KD树性能迅速下降^[4],搜索复杂度接近于线性,面临“维度灾难”问题.此外,树形索引结构所需存储空间大,很难适应大规模数据的存储需求.语义哈希索引算法主要以局部敏感哈希(locality sensitive hashing,简称LSH)为代表,其基本思想是利用多个哈希函数把高维空间中的向量映射到低维空间.然而,基于LSH的索引算法大多采用随机映射,导致索引结果不可控,随着编码位数的上升,索引结果提升缓慢^[5].近年来,有很多研究人员在LSH基础上,通过引进机器学习算法学习最优哈希函数,构造紧凑的二进制编码来解决随机映射产生的不足.

本文针对高维数据索引技术存在的问题,提出了一种监督式的哈希函数学习算法,利用学习到的哈希函数将原始空间的特征向量映射为二进制编码,通过计算编码的汉明距离进行检索.本文的主要目的是针对目前编码方式判别性不足的缺点,利用机器学习算法学习哈希函数,使得量化后的编码不但能够增强判别性,而且能够在一定程度上保证相同语义特征的局部结构性.本文提出的方法充分利用了数据的标签信息,首先,通过判别式学习算法得到子空间,再利用投影后的特征学习偏移参数,得到哈希函数.由于使用判别式的学习方式,得到的每一位汉明编码对数据集中的某一类具有较强的区分能力.本文的主要贡献如下:

- 1) 充分利用了数据的标签信息,通过判别式的算法学习子空间,同时考虑数据之间的局部性和判别性;
- 2) 用相对稀疏的方式进行编码,每个哈希函数只对数据集中的某一类具有很强的区分能力;
- 3) 生成的二进制编码具有较高的判别能力,同时在一程度上保证相同语义特征的局部性.

最后,本文在MNIST和CIFAR-10两个数据集上进行了实验比较和分析,结果表明:对于同样长度的编码,本文提出的方法得到了更高的准确率;同时,也能保证较好的查全率.此外,本文提出的编码方法不但能够获得较高的判别性,同时也能在很大程度上保持相同语义特征的局部性.

本文第1节对现有工作进行介绍.第2节详细叙述本文提出的哈希函数学习算法.第3节通过实验对提出的算法进行验证和比较,对实验结果进行讨论和分析.第4节对本文进行总结,并提出后续研究方向.

1 相关工作

针对大规模数据设计快速索引算法具有广泛的应用价值,例如目标识别^[6,7]、图像检索^[8,9]、图像匹配^[10,11]等.在构建高效的大规模数据检索系统时,存在两个最主要的瓶颈^[12]:数据的存储需求和检索效率.图像、视频等数据往往具有高维度的特征,如果不对原始数据进行压缩,则需要非常大的空间来存储,随着时间的推移,数据量以飞快的速度增长,这样的存储方式将越来越难以适应.同时,数据规模大必将对检索造成很大的不便.显然,在这种情况下,线性的检索方式已经不合适.针对这两个主要的问题,近年来提出了很多方法,它们将原始数据进行压缩,用二进制编码表征原始特征,从而减少存储需求;同时,通过编码对原始数据构建合理的索引,通过索引可以实现快速高效的检索.

树形索引结构是最早得到应用的高维数据索引算法.使用最广泛的传统树结构索引是以KD树^[3]为基础的维度分割方法.它根据特征向量各维度的数据分布在相应层进行剪枝筛选,在低维空间中能够有效地降低搜索复杂度.但Gionis等人^[4]指出:当向量维度超过20时,KD树性能迅速下降,搜索复杂度接近于线性,面临“维度灾难”问题.Beis等人^[13]引入一种停止策略来提高搜索效率,提出了一种类似于KD树的索引结构.浙江大学

Zhuang 等人^[14]提出了基于对称编码的混合距离树索引结构,将近邻搜索从高维空间投影到一维空间.然而,树形索引结构所需存储空间较大,很难适应大规模数据的存储需求,这决定了其不适用目前大规模图像数据集的快速检索.

早期的语义哈希索引算法主要以 LSH 为代表,利用随机产生的哈希函数对原始特征进行索引,达到降低维度的目的.通过若干个哈希函数的组合,将 M 维的原始特征索引成 N 维,其中, $N \ll M$. 随机产生的哈希函数 H 满足以下条件:对于高维空间中的两个数据点 p 和 q ,给定一个阈值 R 和近邻因子 $c > 1$,如果 p 和 q 距离小于 R ,那么 $\Pr(H(p) == H(q)) \geq P1$,即原始空间近邻在通过哈希函数映射后,其值相等的概率至少为 $P1$;反之,如果 p 和 q 的距离大于 cR ,那么 $\Pr(H(p) == H(q)) \leq P2$,即原始空间非近邻在通过哈希函数映射后,其值相等的概率最多为 $P2$.如果满足 $P1 > P2$,这组函数 H 被称为 $(R, cR, P1, P2)$ 敏感哈希函数^[4].在原始 LSH 的基础上, Datar 等人^[15]提出了 E2LSH,基于 p -stable 分布, E2LSH 随机产生投影向量 a ,在区间 $[0, r]$ 内随机产生偏移量 b ,对原始特征 v ,利用函数 $h_{a,b} = \lfloor (a \cdot v + b) / r \rfloor$ 进行降维.类似的工作有普林斯顿大学的 Lü 等人^[16]提出的 Multi-probe LSH,其索引构造方式与 E2LSH 类似,但将特征同时映射到一个哈希表的多个桶来保证查全率.由于其随机性,LSH 相关方法大多数都不依赖于数据本身.然而,随着量化后编码位数的增长,基于 LSH 的方法理论上使得原始空间的度量渐进式的保存在汉明空间^[12],因此,这些方法通常需要更长的编码才能达到令人满意的准确率,而长编码容易导致查全率的降低.此外,这类方法随机产生哈希函数,导致索引结果不可控,随着索引编码位数的上升,索引结果提升缓慢.

近年来,研究者针对 LSH 相关方法存在的问题,利用数据本身的一些性质,例如局部性,提出了一系列基于机器学习的哈希函数学习算法,利用哈希函数将高维特征映射为紧凑的二进制编码,在较低的时间复杂度内完成搜索.为了得到这样的编码,研究者们提出了很多算法,根据学习机制,大致可以分为无监督、半监督和有监督这 3 类.在无监督算法中,最具代表性的方法是利用主成分分析(principle component analysis,简称 PCA)进行降维,从而对特征实现索引^[17].相对于随机产生哈希函数的方法而言,这类依赖数据本身的方法达到了更高的准确率.这类方法还包括谱哈希^[5].谱哈希索引算法对原始高维数据进行谱分析,再通过放松限制条件,将问题转换为拉普拉斯特征图的降维问题从而求解. Liu 等人^[18]提出了一种基于图的索引方法,通过发现原始特征的近邻来学习编码,从而实现了常数时间复杂度的索引.无监督的学习算法大多是以保留原始特征的局部性为优化目标,然而在实际应用中,检索系统往往更倾向于搜索语义相关的结果,因此,标签信息被应用到索引算法中,产生了很多半监督的学习算法. Wang 等人^[19]提出了半监督哈希.他们提出的优化目标是编码的相似性,利用特征对学习,使相似对的二进制编码保持最大程度相似;反之,不相似的特征对编码最大程度不同,同时保证产生的二进制编码在未标记数据上得到的熵最大. Mu 等人^[20]对数据中一些特征对进行人工标注,将其分为语义相似对和语义不相似对,通过放松限制,将优化问题分为若干凸子问题,利用二次规划进行求解,得到哈希函数.这些半监督的方法相对于无监督的方法而言,能够在很大程度上提高二进制编码的语义判别性,具有更实用的价值.此外,还有不少学者提出充分利用数据的标签信息,从而实现了监督式的索引算法. Strecha 等人^[10]首先从数据集中构造一系列相似和不相似的特征对,利用线性判别分析(linear discriminant analysis,简称 LDA),从这些特征对中学习一个最优投影方向,再通过阈值选择得到一个偏移量,将投影方向和偏移量组合形成一个哈希函数,最后,利用哈希函数将特征量化成二进制编码,用于图像匹配.此外,更复杂的方法如 Binary Reconstruction Embeddings(BRE)^[21],以最小化编码的重构误差作为优化目标,提出了一种基于梯度下降的优化算法来学习哈希函数.这类方法达到了较高的准确率,但是其优化过程较为复杂,需要较长的训练时间,因此限制了此类方法在大规模数据集上的适用性^[12].现有的方法还存在一些问题:首先,通过学习哈希函数来获得紧凑的二进制编码,这些投影往往需要较强的判别性,能够同时区分多类特征,否则很难保证检索的准确率,而实际的图像数据往往很难学习到这样的投影,但若增加编码的长度,则需要更多的存储空间,因此,如何在编码位数相同的情况下,使学习到的哈希函数得到更高的检索效果,是一个有待解决的问题.此外,上述方法往往考虑编码之间的局部性或重构误差,存在判别性不足的问题,对于图像检索而言不太适用.

本文针对以上问题,提出了一种监督的判别式哈希函数学习算法,同时考虑了同类数据的局部性和不同类数据的判别性,获得了较好的语义检索性能.此外,与紧凑编码方式不同的是,本文算法采用较为稀疏的编码方

式,但在编码位数相同的情况下,检索性能高于紧凑编码方式.本文通过类似 LDA 的方法学习哈希函数,使得其对某一类数据具有很强的判别性,并保证语义相同特征的局部性.利用学习到的投影向量对原始空间的特征向量进行降维,再学习偏移参数,得到哈希函数.最后,本文在 MNIST 和 CIFAR-10 两个数据集上对提出的方法进行了实验和比较,结果表明:对于同样长度的编码,本文提出的方法能够达到更高的准确率;同时,也能保证较好的查全率,不但提高了编码的判别性,而且能够在很大程度上保持相同语义特征的局部结构.

2 哈希函数学习算法

给定一组数据集 $X=\{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^d$, 其中 x_i 表示 d 维特征向量.索引的目的是找到一组适当的哈希函数 $h: \mathbb{R}^d \mapsto \{1, 0\}^1$, 其中,每个哈希函数将一个特征向量映射成一位哈希值.将一组哈希函数组合,构成一个函数组 $G(x)=[h_1(x), h_2(x), \dots, h_g(x)]$, 则 $G: \mathbb{R}^d \mapsto \{1, 0\}^g$ 将一个特征从原始空间映射到低维空间的编码,其中, $g \ll d$.低维空间编码可以有效减少数据量,实现高效的存储.同时,我们更希望找到一个编码方式,可以区分不同类别的特征.此外,编码的局部性保留能力,即在原始空间是近邻的特征,在编码空间还是近邻,也是衡量编码优劣的标准之一.显然,这两个标准并不一定同时成立.假设原始空间为欧式空间,用欧式距离衡量特征的相似度,我们有可能找到一组哈希函数,使得在欧式空间中是近邻的特征,在编码空间仍然是近邻;且欧式空间语义相似的特征在索引后,在编码空间具有很高的相似度.然而,一旦欧式距离不能很好地度量特征之间的相似度,即欧式距离小的特征反而不属于同一类(这种现象在实际数据集中非常常见),那么过于注重原始特征局部性的保留,反而会使编码的语义判别性降低,对语义检索而言不适用.

本文提出的监督式学习算法充分利用了标签信息,以编码的判别性为衡量标准,使得特征经过索引后具有很强的类间区分度.我们定义一个哈希函数的形式如下:

$$h(x) = \text{sign}(w^T x + t) \quad (1)$$

其中, $x \in \mathbb{R}^d$ 是原始空间的特征向量, $w \in \mathbb{R}^d$ 是投影向量, t 是偏移参数, $\text{sign}(\cdot)$ 是取符号函数.通过哈希函数,原始空间的特征向量被映射为一位的哈希值,此处以 -1 代替 0.一个哈希函数对某一类具有很强的判别性,换言之,原始特征经过哈希函数索引后,特征集中的某一类将被量化为同一个值,例如 1(0); 而其他类别将被量化为另外一个值,即 0(1).这和紧凑的编码方式有所不同,如图 1 所示.

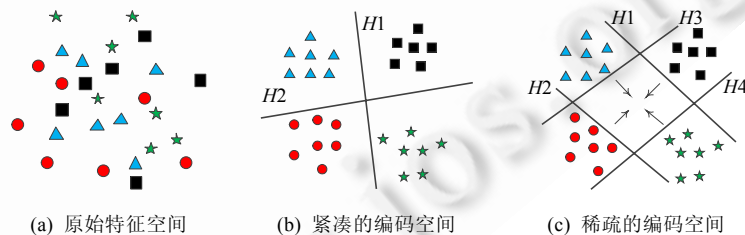


Fig.1 Comparison of code spaces

图 1 编码空间比较

图 1(a)表示原始特征空间,不同形状代表不同类特征.图 1(b)表示紧凑的编码方式,其中, $H1$ 和 $H2$ 是哈希函数,通过两个哈希函数映射,即可将 4 种类别的特征区分,假定 $H1$ 左侧编码为 1,右侧编码为 0, $H2$ 上方编码为 1,下方编码为 0,那么三角形、正方形、圆形、五角星所表示的特征编码依次为 11, 01, 10 和 00.图 1(c)表示稀疏的编码方式,其中, $H1, H2, H3$ 和 $H4$ 表示哈希函数,假定它们朝箭头一侧表示 0,外侧表示 1,那么 4 类特征编码依次为 0010, 1000, 0100 和 0001.不难发现:紧凑的编码有其特有的优势,表示相同类别数的特征,紧凑的编码比稀疏的编码需要更少的位数,这样可以有效地减少存储需求,对于高维数据而言,这个优势尤其明显.然而在语义检索中,紧凑编码对哈希函数判别性的要求则变得更高.从图 1(b)中可以看出:只通过两个哈希函数就将 4 类特征加以区别,这两个哈希函数各自需要对其中两类具有明显的区分效果.然而在实际的数据集中,特别是高维的图

像、视频数据,往往很难找到这样的哈希函数.反之,稀疏的编码方式虽然增加了哈希函数的数量,但提高了其判别性,能够使我们从数据集中更容易学习到区分度高的哈希函数.我们可以推断出:如果哈希函数的判别能力不够,即使使用紧凑的编码方式,仍然需要增加编码位数来达到令人满意的检索效果.换言之,如果在稀疏的编码方式中能够学习到高判别性的哈希函数,则同样可以减少编码的数量.本文的目的就是通过学习具有较强判别能力的哈希函数,利用稀疏的编码方式达到较好的检索效果.学习哈希函数的步骤主要分为两步: W 步骤和 T 步骤,分别学习投影向量 w 和偏移参数 t .

2.1 W 步骤:投影向量学习

假设 $X=\{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^d$ 为给定的零均值 d 维特征向量集合, $Y=\{y_1, y_2, \dots, y_N\}$ 为数据标签, $H=\{h_1, h_2, \dots, h_K\}$ 为一组哈希函数,那么任意一个特征向量经过第 k 个哈希函数索引后,得到一位哈希值:

$$h_k(x_i) = \text{sign}(w_k^T x_i + t_k) \quad (2)$$

即,经过一个哈希函数索引后, $y'_i = \text{sign}(w^T x_i + t)$.给定集合 M 和 C , M 是 X 中某一类特征的集合, C 是 M 对于 X 的补集,即 $M \subset X, C \subset X, M \cap C = \emptyset, M \cup C = X$ 且 $\forall (x_i \in M, x_j \in M), y_i = y_j$.构造集合 M 和 C 时,将集合 X 中某一类特征归入 M ,其他特征归入 C .与LDA类似,本文通过衡量 M 和 C 两类的类内散列度(intra-class scatter)和类间散列度(inter-class scatter)来学习哈希函数.LDA假设样本服从一定的分布,用类内样本和类中心距离衡量类内散列度,用类中心之间的距离衡量类间散列度.然而,现实的数据集往往达不到这样的要求,且集合 C 中包含多种类别的样本,因此,简单地使用LDA并不合适.

考虑到最终通过哈希函数映射得到的编码能够获得较好的判别性,且保持 M 类特征在原始空间的局部性,本文采取以下策略,将子空间学习描述为二分类问题:

- (1) 对于集合 M 中的样本,通过计算近邻之间的距离来衡量类内散列度;
- (2) 对于集合 C 中的样本,通过计算样本与样本中心的距离来衡量类内散列度;
- (3) 通过计算 M 和 C 样本中心的距离来衡量类间离散度.

对于策略(1),由于集合 M 中属于同类样本,因此考虑样本之间的局部性,使得在原始特征空间是同类且是近邻的特征,经过哈希函数索引后,它们之间的距离能够最小.对于策略(2),由于集合 C 包含多类样本,如果考虑局部性,反而会使原始空间中异类但却是近邻的样本在索引后保持近邻关系,因此,策略(2)不考虑特征之间的局部性.策略(3)与LDA一致,通过样本中心距离最大化来区分两类样本.

定义损失函数 \mathcal{L} :

$$\mathcal{L}(w, t) = \frac{\mathcal{L}_B}{\mathcal{L}_W} = \frac{\mathcal{L}_B}{\mathcal{L}_M + \mathcal{L}_C} \quad (3)$$

其中, \mathcal{L}_B 为类间散列度, \mathcal{L}_W 为类内散列度.那么,根据上述3个策略可以分别得到:

$$\mathcal{L}_M = \sum_{x_i \in M} \sum_{x_j \in NN(x_i)} \Phi(h, x_i, x_j) \quad (4)$$

$$\mathcal{L}_C = \sum_{x_i \in C} \Phi(h, x_i, \mu_C) \quad (5)$$

$$\mathcal{L}_B = \Phi(h, \mu_C, \mu_M) \quad (6)$$

其中, $NN(x_i)$ 表示特征 x_i 的近邻集合, $\Phi(h, x_i, x_j)$ 表示特征 x_i 和 x_j 经过哈希函数 h 索引后的差异, μ_C 和 μ_M 分别是集合 M, C 的样本中心. $\Phi(h, x_i, x_j)$ 可表示为

$$\Phi(h, x_i, x_j) = \|y'_i - y'_j\|^2 \quad (7)$$

即,衡量特征 x_i 和 x_j 对应哈希值 y'_i 和 y'_j 的差异:若 $y'_i = y'_j$,则 $\Phi(h, x_i, x_j)$ 取较小值;若 $y'_i \neq y'_j$,则 $\Phi(h, x_i, x_j)$ 取较大值.因此可以看出,求得最优的哈希函数即优化目标函数 $\mathcal{L}(w, t)$:

$$(w^*, t^*) = \arg \max_{w, t} \mathcal{L} \quad (8)$$

进一步来说, \mathcal{L}_M 可表示为

$$\mathcal{L}_M = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} \|y'_i - y'_j\|^2 = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} \|\text{sign}(w^T x_i + t) - \text{sign}(w^T x_j + t)\|^2 \quad (9)$$

由于函数 $h(x)$ 中有不可微的取符号函数,因此直接优化公式(8)非常困难.文献[5,10]针对这个问题采取谱松弛(spectral relaxation)的策略,将取符号函数去除再进行优化,这样的近似优化方法被证明具有较好的效果.因此,本文在优化目标函数(8)时,放松对 $\text{sign}(\cdot)$ 函数的限制,将 $\text{sign}(\cdot)$ 函数去掉,得到新的目标函数 $\tilde{\mathcal{L}}_M$:

$$\tilde{\mathcal{L}}_M(w, t) = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} \|w^T x_i - w^T x_j\|^2 \quad (10)$$

在本文的问题中,这样的放松限制有两个好处:首先,将不可微的 $\text{sign}(\cdot)$ 函数去掉,有利于优化目标函数;其次,求取近似解时可以暂时抵消偏移量 t 的影响,从而将求解公式(8)转换为先学习投影向量 w ,再学习偏移量 t 这两个步骤,从而求得原问题的近似解.同样地,对于任意 $x_i \in C$,有:

$$\tilde{\mathcal{L}}_C(w, t) = \sum_{x_i \in C} \Phi(h, x_i, \mu_C) = \sum_{x_i \in C} \|w^T x_i - w^T \mu_C\|^2 \quad (11)$$

$\tilde{\mathcal{L}}_B$ 可表示为

$$\tilde{\mathcal{L}}_B(w, t) = \Phi(h, \mu_C, \mu_M) = \|w^T \mu_C - w^T \mu_M\|^2 \quad (12)$$

对于 $\tilde{\mathcal{L}}_M$, 可以得到如下的表示:

$$\tilde{\mathcal{L}}_M(w, t) = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} \|w^T x_i - w^T x_j\|^2 = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} w^T (x_i - x_j)(x_i - x_j)^T w = w^T \tilde{\mathcal{L}}'_M w \quad (13)$$

其中,

$$\tilde{\mathcal{L}}'_M = \sum_{x_i \in \mathcal{M}} \sum_{x_j \in NN(x_i)} (x_i - x_j)(x_i - x_j)^T \quad (14)$$

同样地,我们可以得到:

$$\tilde{\mathcal{L}}'_C = \sum_{x_i \in C} (x_i - \mu_C)(x_i - \mu_C)^T \quad (15)$$

$$\tilde{\mathcal{L}}'_B = (\mu_C - \mu_M)(\mu_C - \mu_M)^T \quad (16)$$

最后,目标函数可表示为

$$(w^*, t^*) = \arg \max_{w, t} \tilde{\mathcal{L}} = \arg \max_{w, t} \frac{w^T \tilde{\mathcal{L}}'_B w}{w^T (\tilde{\mathcal{L}}'_M + \tilde{\mathcal{L}}'_C) w} \quad (17)$$

其中, $\tilde{\mathcal{L}}'_M$, $\tilde{\mathcal{L}}'_C$ 和 $\tilde{\mathcal{L}}'_B$ 都可以从训练集得到.目标函数的求解过程与 LDA 相似,在引入拉格朗日乘子 λ 后,我们可以得到:

$$(\tilde{\mathcal{L}}'_M + \tilde{\mathcal{L}}'_C)^{-1} \tilde{\mathcal{L}}'_B w = \lambda w \quad (18)$$

w 即矩阵 $(\tilde{\mathcal{L}}'_M + \tilde{\mathcal{L}}'_C)^{-1} \tilde{\mathcal{L}}'_B$ 最大的特征值对应的特征向量,其中每一个特征向量可以表示一个哈希函数的投影向量.假设我们取特征值最大的 P 个特征向量,得到投影矩阵 W' ,其对应了 P 个哈希函数.

2.2 T步骤:偏移量学习

在 W 步骤学习得到了投影向量 w 之后,我们在 T 步骤学习哈希函数 $h(x_i) = \text{sign}(w^T x_i + t)$ 中的偏移参数 t .假设矩阵 W' 是 W 步骤学习到的投影矩阵,由于原问题比较复杂,在 W 步骤进行了限制放松.因此,通过 W 步骤和 T 步骤进行优化,是希望得到原问题的一个近似解.在 T 步骤,我们重新考虑取符号函数 $\text{sign}(\cdot)$ 的影响,这可以看作是对去掉取符号函数的一种补偿.因此,在 T 步骤我们通过如下的目标函数进行偏移量的学习:

$$L(T) = L_1(T) - L_2(T) \quad (19)$$

$$L_1(T) = \frac{1}{m} \sum_{(x_i, x_j) \in \mathcal{M}} \text{sign}(W'x_i + T)^T \text{sign}(W'x_j + T) \quad (20)$$

$$L_2(T) = \frac{1}{c} \sum_{x_i \in \mathcal{M}, x_j \in C} \text{sign}(W'x_i + T)^T \text{sign}(W'x_j + T) \quad (21)$$

其中, T 表示 W' 中每个哈希函数对应的偏移量组成的向量.

从 M 中随机采样得到相同语义的样本 m 对, 分别从 M 和 C 中随机采样得到不同语义的样本 c 对. $L_1(T)$ 可进一步表示为

$$L_1(T) = \frac{1}{m} \sum_{(x_i, x_j) \in M} \sum_{p=1}^P (\text{sign}(W_p^T x_i + t_p) \text{sign}(W_p^T x_j + t_p)) \quad (22)$$

其中, W_p 表示 W' 中对应 W 步骤的第 p 个特征向量, t_p 表示对应的第 p 个偏移量.

我们可以发现, W' 和 T 的每一个分量在计算过程中不会相互影响. 因此在 W 步骤中, 我们采取将参数 T 的每一个分量分开求取的策略, 每次求取一个分量. 因此, 求取偏移量 T 的目标函数可表示为

$$t_p^* = \arg \max_{t_p} \left(\frac{1}{m} \sum_{(x_i, x_j) \in M} \text{sign}(W_p^T x_i + t_p) \text{sign}(W_p^T x_j + t_p) - \frac{1}{c} \sum_{x_i \in M, x_j \in C} \text{sign}(W_p^T x_i + t_p) \text{sign}(W_p^T x_j + t_p) \right) \quad (23)$$

由于在 W 步骤中已经得到了投影矩阵 W' , 因此在上述优化公式中, 只有 1 个未知参数 t 需要求解. 令

$$f(t) = \text{sign}(W_p^T x_i + t) \text{sign}(W_p^T x_j + t) = \text{sign}((W_p^T x_i + t)(W_p^T x_j + t)) = \text{sign}(g(t)) \quad (24)$$

其中, 由于 $a_i = W_p^T x_i$ 和 $a_j = W_p^T x_j$ 是已知的, 因此, $g(t)$ 是关于 t 的二次函数: $g(t) = t^2 + (a_i + a_j)t + a_i a_j$. 那么,

- 当 $t < \min(-a_i, -a_j)$ 或 $t > \max(-a_i, -a_j)$ 时, $f(t) = 1$;
- 当 $\min(-a_i, -a_j) < t < \max(-a_i, -a_j)$ 时, $f(t) = -1$.

因此, 对参数 t_p^* 的目标函数可以表示为

$$t_p^* = \arg \max_{t_p} \left(\frac{1}{m} \sum_{(x_i, x_j) \in M} F(t_p < \min(-a_i, -a_j) \parallel t_p > \max(-a_i, -a_j)) + \frac{1}{c} \sum_{x_i \in M, x_j \in C} F(\min(-a_i, -a_j) < t_p < \max(-a_i, -a_j)) \right) \quad (25)$$

其中, $F(x)$ 衡量数据集满足条件 x 的程度.

在本文中使用时频统计, 即给定 t 时, $\sum_{(x_i, x_j) \in M} F(t < \min(-a_i, -a_j) \parallel t > \max(-a_i, -a_j))$ 表示在集合 M 满足条件 $t < \min(-a_i, -a_j) \parallel t > \max(-a_i, -a_j)$ 的特征对 (x_i, x_j) 的总数. 由于目标函数中只有 1 个未知参数 t , 因此通过简单的线性搜索, 即可得到最优的参数 t . T 步骤将每一个特征向量和对应的偏移量 T 进行分离求解, 每一步求得 T 的一个分量, 经过 P 次求解后, 最终求得偏移参数 T .

在实际应用中, 如果需要的编码长度不大 (例如 64 位), 则可以在 W 步骤取一个特征向量, 在 T 步骤对应学习一个偏移量, 同时对数据集进行多次随机采样, 针对同一类特征重复学习多个哈希函数, 对每类特征的哈希函数根据其二值化效果进行排名, 选取最好的若干个哈希函数. 如果需要的编码长度较大 (如 256 位), 则为了避免每次训练一个哈希函数耗费时间, 可以在 W 步骤中一次选取多个特征向量, 再按照本节的方法分离训练对应的偏移量 T .

3 实验分析

为了对本文提出的算法进行验证和分析, 我们从已经发表的文献中选取了最具代表性的方法进行比较, 包括 AGH^[18], KSH^[12], LDAH^[10], AQBC^[22], BRE^[21], E2LSH^[15], ITQ^[23], SH^[5] 和 SIKH^[17]. 本文提出的算法在实验中用 DH 表示. 对于 AGH 方法, 其参数设置与文献[18]相同, MNIST 检索参数设为 $m=300, s=2$; 对 CIFAR-10 设为 $m=300, s=5$. BRE 方法的参数设置与文献[21]相同, 采用监督式设置, 用同类样本代替近邻进行训练. SIKH 方法的参数设置与文献[17]相同, 使用的核参数 $\gamma=1$. 本文选取了在图像数据索引领域常用的两个数据集 MNIST 和 CIFAR-10 进行实验, 分别对编码的判别能力和编码对特征在原始空间局部性的保留能力进行了比较, 同时还对各算法量化的时间进行了比较. 实验的计算机环境是: 处理器为 Intel i5-2430M, 主频 2.4GHz, 4 核; 内存 4GB; 硬盘 500GB; 操作系统为 Windows 7 Ultimate 64 位; 编程环境为 Matlab R2011a.

3.1 实验设置

3.1.1 数据集

本文进行实验的数据集有两个,分别是 MNIST 和 CIFAR-10.这两个数据集在图像数据检索中是常用的两个标准数据集^[12,23,24].MNIST 是著名的手写数字数据库,由谷歌实验室的 Corinna Cortes 和 Courant 研究所的 YannLeCun 建立,数据库包含数字 0~9 的手写体,以像素数据存储.其中,训练集图像有 60 000 张,测试集图像有 10 000 张.在本文的实验中,将所有 0~9 的数据还原成图像,并对图像提取 512 维的全局 GIST 特征,将每个数字对应的图像分组,每组分别取出 300 张图像作为训练集学习哈希函数,每组分别取出 300 张图像作为查询用例,剩余的图像作为数据库进行索引.CIFAR-10 数据库是 80 million tiny image^[6]数据集中已被标记的一个子数据集.CIFAR-10 数据集共有 60 000 幅 32×32 的彩色图像,共分为 10 个类,每类 6 000 幅.在本文的实验中,对每幅图像提取 512 维的全局 GIST 特征,将每类图像进行分组,每组取出 300 张图像作为训练集学习哈希函数,每组分别取出 300 张图像作为查询用例,剩余图像作为数据库进行索引.为了表述方便,对于上述两个数据集,用 T 表示训练集, D 表示数据库, Q 表示查询用例.

3.1.2 评价标准

本文在实验中主要对算法的 3 个指标进行了比较,分别是编码的判别能力、局部性保留能力和编码时间.前两种性质均用准确率和查全率衡量,编码时间指索引算法把数据库量化为编码耗费的平均时间.在实验中,每个算法对数据集进行了 10 次量化,取平均的编码时间作为结果,时间单位为 s.

编码的判别能力是指:特征 A 和 B 在原始空间有相同的语义,即属于同一类特征,那么特征 A 和 B 在经过映射后,在编码空间中的相似度大,在实验中表现为汉明距离小.具体来说,假设特征 A 在原始空间有相同语义的特征集合为 $Sem(A)$,其元素个数为 T ,在经过哈希函数索引后,以汉明距离 $dist$ 为阈值进行检索,将汉明距离小于 $dist$ 的特征集合 $Sem(H(A))$ 作为检索结果,其元素个数为 G , $S=Sem(A) \cap Sem(H(A))$ 表示检索结果中与特征 A 语义相同的元素集合,其元素个数为 H .

在判别性实验中,准确率可表示为 $Precision=H/G$,查全率可表示为 $Recall=H/T$.

编码的局部性保留能力是指:特征 A 和 B 在原始空间(如欧式空间)的度量方式 Dis (欧式距离)中是近邻,那么,特征 A 和 B 在经过映射后,在编码空间(如汉明空间)的度量方式 Dis' (汉明距离)中仍然是近邻.具体来说,假设特征 A 在原始空间的 K 近邻为集合 $NN(A)$,其元素个数为 T ,在经过哈希函数索引后,以汉明距离 $dist$ 为阈值进行检索,将汉明距离小于 $dist$ 的特征集合 $NN(H(A))$ 作为检索结果,其元素个数为 G , $S=NN(A) \cap NN(H(A))$ 表示检索结果中在原始空间是特征 A 近邻的元素集合,其元素个数为 H .

在局部性保留实验中,准确率可表示为 $Precision=H/G$,查全率可表示为 $Recall=H/T$.

局部性保留的强弱可以衡量哈希函数对原始特征的简化能力,但值得注意的是:如果原始空间的度量方式不能很好地表达判别性,即原始空间中距离小的特征反而不属于同一类(这种现象在实际数据集中非常常见),那么,编码过程过度地注重局部性保留反而会对检索产生负面影响.因此,本文实验从这两方面进行分析,希望在获得令人满意的判别能力的同时,能够最大程度地兼顾局部性保留.

3.2 判别性实验与分析

在判别性实验中,用 T 部分数据进行学习得到哈希函数.由于特征维数为 512,编码位数一般不会超过特征维数,因此对 Q 和 D 部分数据分别量化为 64,128,256 位汉明编码.在检索中以准确率和查全率为标准,在查询过程中汉明距离依次增加,得到如图 2、图 3 所示的 P - R 曲线.

从图 2 中可以看出:本文提出的方法在 MNIST 和 CIFAR-10 上均取得了对比算法更好的效果,尤其是在 MNIST 数据集上.其主要原因可以总结为:AGH 这类基于无监督哈希函数学习的索引算法通常考虑原始特征的局部性,可以在编码空间很好地保留近邻关系;然而,由于原始空间的度量方式往往缺乏语义判别能力,因此在语义检索中效果不理想.BRE 虽然采用监督的学习算法,但它以最小化编码对原始特征的重构误差为目标,没有体现编码的判别能力.传统的 E2LSH 则采用随机投影的策略,只能保证在一定的概率范围内使相似的原始特

征具有相同的编码,因此同样没有考虑判别性.本文方法是针对稀疏的编码方式设计的,与采用 LDA 方法的 LDAH 相比具有明显的区别:首先,本文只考虑样本中某一类的局部性,用近邻之间的散列度来描述,与 LDA 采取的类内散列度是不同的,后者描述的是每类样本到其样本中心的距离;其次,本文考虑最大化某一类和剩余类别样本的类间散列度,而 LDA 最大化样本中心之间的距离,实验结果表明,本文设计的方法取得了更好的效果.本文提出的索引算法采用有监督的哈希函数学习,首先,考虑不同语义特征之间的判别性,使其在索引后最大化,因此,在语义检索中能够有效区分不同语义的特征;其次,算法考虑了相同语义特征之间的局部性,在索引后,使相同语义的特征保持近邻关系.这两点促使最终语义检索的结果优于其他对比算法.

从图 3 中可以看出:对于数据集 CIFAR-10,所有方法的实验结果均有比较明显的下降,原因在于 CIFAR-10 图像本身比较复杂,但在相同测试环境下,本文提出的方法均优于其他算法.其主要原因可以总结为:AGH 这类基于无监督哈希函数学习的索引算法和基于随机投影的 E2LSH 算法,其实验结果与在 MNIST 上的结果相比下降尤为明显,这是因为 MNIST 数据集类间差异大于类内差异,且图像本身比较简单,在原始特征空间中,近邻之间的语义大部分是相似的,因此以保持近邻关系为优化目标,也可以得到较好的判别性.然而在 CIFAR-10 数据集上,类内差异和类间差异区别没有 MNIST 数据集那么显著,近邻之间的语义往往是不同的,因此,考虑局部性反而导致在索引后判别性下降得非常严重.而 BRE 这类监督的学习算法因为考虑了原始特征的语义,在 CIFAR-10 上的测试结果相对于其他无监督的方法而言下降幅度比较小.由于 CIFAR-10 数据集样本差异不明显,LDAH 算法很难学习到一个可以很好地同时区分多类的子空间,因此实验效果不佳.本文提出的算法充分考虑了判别性,因此在 CIFAR-10 数据集上表现出优于其他对比算法的性能.

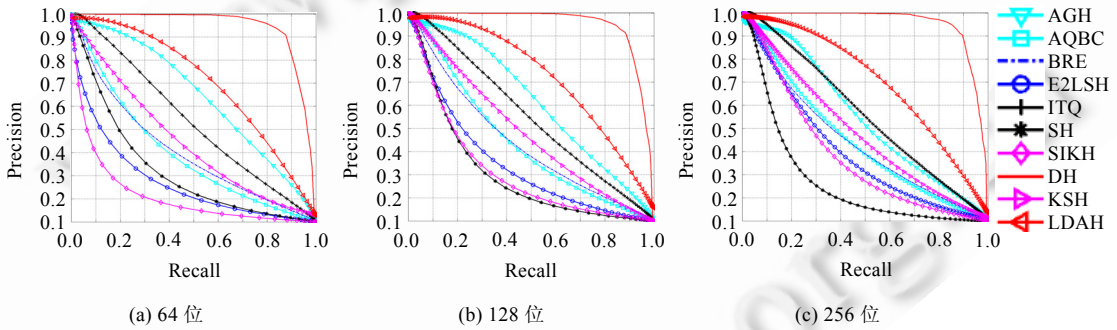


Fig.2 Discriminative test on MNIST

图 2 MNIST 判别性测试

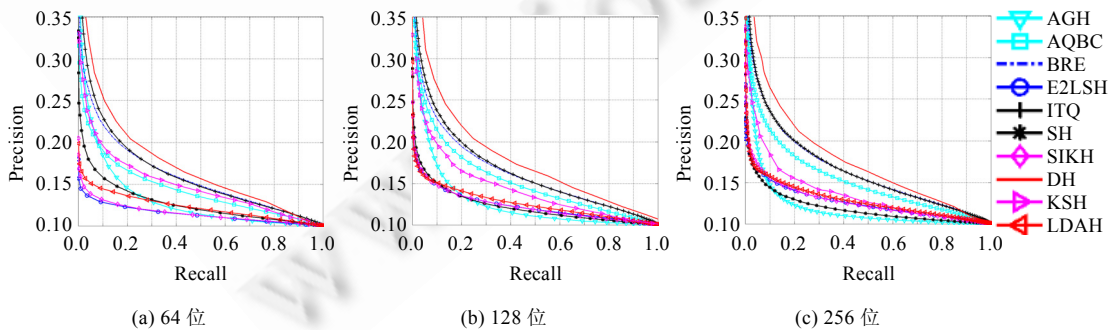


Fig.3 Discriminative test on CIFAR-10

图 3 CIFAR-10 判别性测试

综合上述分析可以得出以下结论:本文提出的方法考虑了不同语义特征之间的判别性和相同语义特征之间的局部性,在两个数据集上进行的语义检索实验中都取得了对比算法更好的检索性能.可以看出:本文提出

的判别式学习算法学习到的哈希函数在语义检索中是十分有效的,相对于其他方法而言,可以更好地区分不同的类别.对于相同长度的编码,由于本文算法提高了每一位编码的判别性,使用稀疏的编码方式依然得到了比紧凑编码方式更好的检索效果.

3.3 局部性保留实验与分析

在局部性保留实验中,首先对 Q 部分数据的每一个样本,其原始特征(实验中用欧氏距离)和 D 数据的原始特征进行近邻搜索, Q 中每个样本得到 2 000 个近邻,作为最后检索结果的评价标准.用 T 部分数据进行学习得到哈希函数,把 Q 和 D 数据量化为 64,128 和 256 位的编码,以原始空间的近邻为检索标准得到如图 4、图 5 所示的 $P-R$ 曲线.

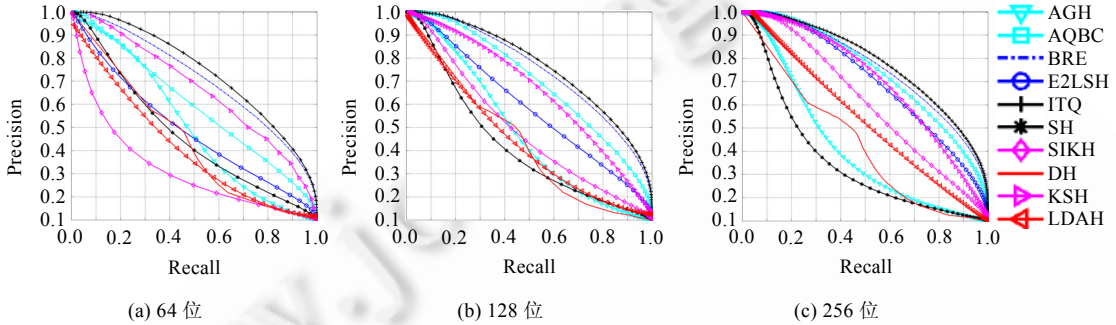


Fig.4 Locality preserving test on MNIST

图 4 MNIST 局部保留性测试

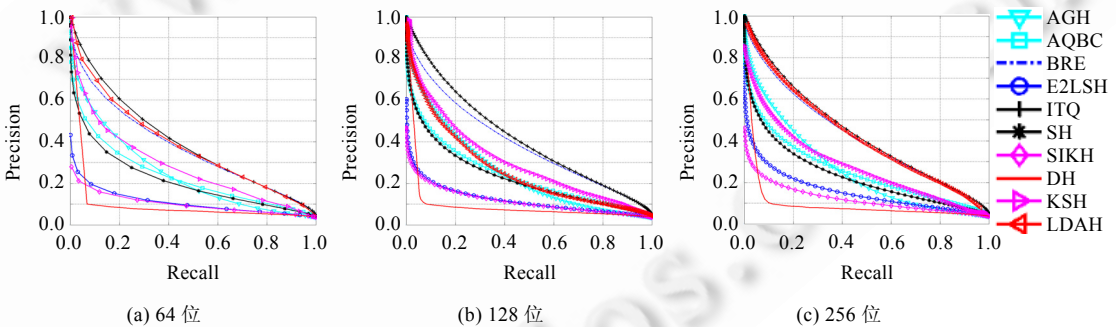


Fig.5 Locality preserving test on CIFAR-10

图 5 CIFAR-10 局部保留性测试

本文的方法与无监督式的学习算法相比差别比较明显.由于本文方法的目的在于图像语义检索,考虑提高编码的判别性来区分不同语义的特征,因此对于语义不同的特征,局部性保留能力并不是特别重要,对于语义相同的特征,编码以后仍然需要在一定程度上保证其近邻关系,这与提出的哈希函数学习算法是一致的.即在投影子空间学习过程中考虑相同语义特征的局部性、不同语义特征的判别性.

编码后,局部性的下降在一定程度上是由判别性的提高造成的,分析如下:

- (1) 对于原始特征判别能力较强的数据集,在提高判别性后,编码的局部性会下降,但下降程度不会很大.其原因在于:由于编码提高了判别性,因此原始特征中语义不同的近邻在编码后变成了非近邻,原始特征中语义相同的近邻在编码后有较大可能仍然是近邻.
- (2) 对于原始特征判别能力较弱的数据集,编码的局部性会有明显下降.其原因在于:原始空间中的近邻有相当一部分语义不同(原始特征判别能力弱),在提高编码的判别性后,这部分近邻变成了非近邻.

图 4 和图 5 很好地说明了这两种情况:对于图 4,由于 MNIST 原始特征类间差异比较明显,因此本文算法在编码后,局部性有所下降,但下降程度不是很大;对于图 5,由于 CIFAR-10 原始特征类间差异和类内差异区别不明显,因此在提高判别性后,局部性保留能力必定会有明显的下降.这两点从侧面说明:对于语义检索而言,本文算法可以有效地增强编码的判别能力.

3.4 算法量化时间比较实验

在量化时间的比较实验中, T 部分数据包括 3 000 个样本,共 10 类,每类 300 个, D 数据在两个数据集上分别取 50 000 和 54 000,我们将各种算法在两个数据集上对数据的 D 部分进行量化,分别量化为 64,128,256 位编码,每个算法在一个数据集上量化 10 次,取平均值作为量化时间.时间比较的结果见表 1 和表 2.本文用于对比的算法大部分是基于无监督学习的算法,因此在时间上显然要快于基于监督学习的算法.其中,BRE 由于需要不断迭代,训练时间最长.本文提出的算法通过监督学习增强了编码的判别性,虽然以小部分训练时间为代价,但是显著提高了语义检索性能.此外,与监督式的 BRE 算法相比,本文的学习过程相对简单,因此耗时更少.

最后,图 6 给出了应用本文算法在两个实验数据集上的检索结果示例,其中,左列为查询图像,右边每行对应的是检索结果.图中矩形框表示检索结果中与查询图像语义不相似的图像.

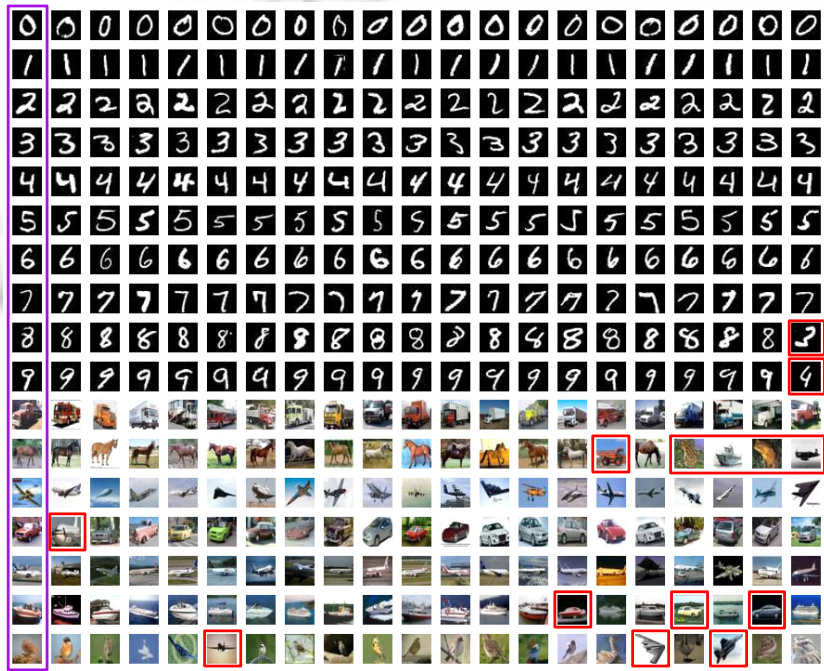


Fig.6 Examples of retrieval results

图 6 查询结果示例

Table 1 Coding time comparison on MNIST

表 1 MNIST 编码时间比较

编码时间(s)	AGH	AQBC	BRE	E2LSH	ITQ	SH	SIKH	DH	KSH	LDAH
64 位	3.1	8.1	1 478.3	12.7	5.2	2.1	1.9	300.1	420.6	4.0
128 位	4.3	13.3	3 103.5	18.6	10.4	4.1	2.2	399.2	830.5	4.7
256 位	5.5	25.6	5 846.3	28.1	28.8	7.7	3.4	470.5	1 700.2	5.3

Table 2 Coding time comparison on CIFAR-10**表 2** CIFAR-10 编码时间比较

编码时间(s)	AGH	AQBC	BRE	E2LSH	ITQ	SH	SIKH	DH	KSH	LDAH
64 位	3.3	8.2	1 610.3	14.5	5.5	2.2	1.6	310.0	440.3	4.3
128 位	4.2	12.9	3 233.0	17.4	10.6	4.0	2.5	410.5	890.4	4.7
256 位	6.0	25.5	5 799.8	29.9	30.1	6.1	3.1	500.3	1 780.1	5.1

4 结束语

本文考虑编码对原始数据的判别性和局部性保留,提出了一种基于子空间学习的图像语义哈希索引方法.考虑同类特征之间的局部性和不同类特征之间的判别性,通过放松限制学习哈希函数的投影向量,通过统计的方法学习哈希函数的偏移量.实验结果表明该方法是有效的.然而,本文的方法还存在一些问题有待解决:首先,本文采取的编码方式属于稀疏的编码方式,虽然通过提高每个哈希函数的判别性,使得在编码长度相同时,与紧凑的编码方式相比具有更好的检索效果,但随着数据量的增加和数据类间差异的减小,编码效果可能会下降.

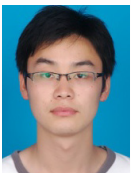
未来的研究工作可以致力于以下两个方面:

- 1) 在本文工作的基础上,将稀疏的编码方式扩展为紧凑的编码方式,同时保证每位编码仍然具有较高的判别性;
- 2) 对类间差异较小的数据,需要研究判别性更好的算法来学习哈希函数.

References:

- [1] <http://venturebeat.com/2008/07/25/google-finds-that-the-web-has-over-1-trillion-unique-urls/>
- [2] <http://www.kullin.net/2010/09/flickr-5-billion-photos/>
- [3] Arya S, Mount DM. Approximate nearest neighbor queries in fixed dimensions. In: Proc. of the 4th Annual ACM/SIGACT-SIAM Symp. on Discrete Algorithms. New York: ACM/SIAM, 1993. 271–280.
- [4] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing. In: Proc. of the 25th Int'l Conf. on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers, 1999. 518–529.
- [5] Weiss Y, Torralba A, Fergus R. Spectral hashing. In: Proc. of the 22th Annual Conf. on Neural Information Processing System. New York: Curran Associates Inc., 2008. 1753–1760.
- [6] Torralba A, Fergus R, Freeman WT. 80 million tiny images: A large dataset for non-parametric object and scene recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008,30(11):1958–1970. [doi: 10.1109/TPAMI.2008.128]
- [7] Torralba A, Fergus R, Weiss Y. Small codes and large databases for recognition. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587633]
- [8] Kulis B, Jain P, Grauman K. Fast similarity search for learned metric. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(12):2143–2157. [doi: 10.1109/TPAMI.2009.151]
- [9] Xu H, Wang JD, Li Z, Zeng G, Li SP, Yu NH. Complementary hashing for approximate nearest neighbor search. In: Proc. of the IEEE Int'l Conf. on Computer Vision. New York: IEEE, 2011. 1631–1638. [doi: 10.1109/ICCV.2011.6126424]
- [10] Strecha C, Bronstein AM, Bronstein MM, Fua P. LDAHash: Improving matching with smaller descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(1):66–78. [doi: 10.1109/TPAMI.2011.103]
- [11] Korman S, Avidan S. Coherency sensitive hashing. In: Proc. of the IEEE Int'l Conf. on Computer Vision. New York: IEEE, 2011. 1607–1614. [doi: 10.1109/ICCV.2011.6126421]
- [12] Liu W, Wang J, Ji RR, Jiang YG, Chang SF. Supervised hashing with kernels. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2012. 2074–2081. [doi: 10.1109/CVPR.2012.6247912]
- [13] Beis JS, Lowe DG. Shape indexing using approximate nearest-neighbor search in high-dimensional spaces. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 1997. 1000–1006.

- [14] Zhuang Y, Zhuang YT, Li Q, Chen L, Yu Y. Indexing high-dimensional data in dual distance spaces: A symmetrical encoding approach. In: Proc. of the 11th Int'l Conf. on Extending Database Technology: Advances in Database Technology. New York: ACM, 2008. 241–251. [doi: 10.1145/1353343.1353375]
- [15] Datar M, Immorlica N, Indyk P, Mirrokni VS. Locality sensitive hashing scheme based on p -stable distribution. In: Proc. of the 20th Annual Symp. on Computational Geometry. New York: ACM, 2004. 253–262. [doi: 10.1145/997817.997857]
- [16] Lü Q, Josephson W, Wang Z, Charikar M, Li K. Multi-Probe LSH: Efficient indexing for high-dimensional similarity search. In: Proc. of the 33rd Int'l Conf. on Very Large Data Bases. New York: ACM, 2007. 950–961.
- [17] Raginsky M, Lasebnik S. Locality sensitive binary codes from shift-invariant kernels. In: Proc. of the 23th Annual Conf. on Neural Information Processing System. New York: Curran Associates Inc., 2009. 1509–1517.
- [18] Liu W, Wang J, Kumar S, Chang SF. Hashing with graphs. In: Proc. of the Int'l Conf. on Machine Learning. Madison: Omnipress, 2001. 1–8.
- [19] 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. [doi: 10.1109/TPAMI.2012.48]
- [20] Mu YD, Shen JL, Yan SC. Weakly-Supervised hashing in kernel space. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2010. 3344–3351. [doi: 10.1109/CVPR.2010.5540024]
- [21] Kulis B, Darrell T. Learning to hash with binary reconstructive embeddings. In: Advances in Neural Information Processing Systems 22, Proc. of the 23rd Annual Conf. on Neural Information Processing System. New York: Curran Associates Inc., 2009. 1042–1050.
- [22] Gong YC, Kumar S, Verma V, Lazebanik S. Angular quantization-based binary codes for fast similarity search. In: Advances in Neural Information Processing Systems 25, Proc. of the 26th Annual Conf. on Neural Information Processing System. New York: Curran Associates Inc., 2008. 1205–1213.
- [23] Gong YC, Lazebanik S. Iterative quantization: A procrustean approach to learning binary codes. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2011. 817–824. [doi: 10.1109/CVPR.2011.5995432]
- [24] Norouzi M, Fleet DJ. Minimal loss hashing for compact binary codes. In: Proc. of the Int'l Conf. on Machine Learning. Madison: Omnipress, 2011. 353–360.



毛晓蛟(1988—),男,江苏常熟人,博士生,主要研究领域为图像语义理解,机器学习,海量数据挖掘。
E-mail: xjmgl.nju@gmail.com



杨育彬(1977—),男,博士,副教授,CCF 高级会员,主要研究领域为数字媒体理解,机器学习,海量数据挖掘。
E-mail: yangyubin@nju.edu.cn