

FP-CNNH:一种基于深度卷积神经网络的快速图像哈希算法

刘 冶^{1,2} 潘 炎^{1,3} 夏榕楷² 刘 荻¹ 印 鉴^{1,2}

(中山大学数据科学与计算机学院 广州 510000)¹ (中山大学信息科学与技术学院 广州 510000)²
(中山大学软件学院 广州 510000)³

摘 要 在大数据时代,图像检索技术在大规模数据上的应用是一个热门的研究领域。近年来,大规模图像检索系统中,图像哈希算法由于具备提高图像的检索效率同时减少储存空间的优点而受到广泛的关注。现有的有监督学习哈希算法存在一些问题,主流的有监督的哈希算法需要通过图像特征提取器获取人为构造的图像特征表示,这种做法带来的图像特征损失影响了哈希算法的效果,也不能较好地处理图像数据集中语义的相似性问题。随着深度学习在大规模数据上研究的兴起,一些相关研究尝试通过深度神经网络进行有监督的哈希函数学习,提升了哈希函数的效果,但这类方法需要针对数据集人为设计复杂的深度神经网络,增大了哈希函数设计的难度,而且深度神经网络的训练需要较多的数据和较长的时间,这些问题影响了基于深度学习的哈希算法在大规模数据集上的应用。针对这些问题,提出了一种基于深度卷积神经网络的快速图像哈希算法,该算法通过设计优化问题的求解方法以及使用预训练的大规模深度神经网络,提高了哈希算法的效果,同时明显地缩短了复杂神经网络的训练时间。根据在不同图像数据集上的实验结果分析可知,与现有的基准算法相比,提出的算法在哈希函数训练效果和训练时间上都具有较大的提高。

关键词 深度学习,图像检索,图像哈希,神经网络,优化算法

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.9.007

FP-CNNH: A Fast Image Hashing Algorithm Based on Deep Convolutional Neural Network

LIU Ye^{1,2} PAN Yan^{1,3} XIA Rong-kai² LIU Di¹ YIN Jian^{1,2}

(School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510000, China)¹

(School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510000, China)²

(School of Software, Sun Yat-sen University, Guangzhou 510000, China)³

Abstract In the big data era, application research on image retrieval technology for large-scale data is a hot field. Recent years, image hash algorithm has attracted much attention in large-scale image retrieval system in order to improve the retrieval efficiency and reduce the storage space. However, there are some issues with existing supervised hash code learning algorithms. Most of the supervised hash algorithms need to use image feature extractor for obtaining hand-crafted image features, which influence the effect of hash code training with image features, at the same time these methods cannot deal well with semantic similarity for image data set. With the development of deep learning research on the large-scale data, some recent related work try to deploy deep neural network to learn hash function and make the hash code training effect increased. But such kind of methods require carefully designed complex neural network structure thus increase the difficulty of the hash function design and cost more time on neural network training with large data set. These problems limit the range of the hash algorithm application with the deep learning architecture for large data sets. To solve the problems mentioned above, this paper proposed a fast image hashing algorithm based on deep convolution neural network. The proposed algorithm consists of an optimization approach for constructing the hash code of the training data set and a pre-trained large deep neural network for learning to improve the effect of hash algorithm, shortening the training time of complex neural network. According to the analysis of experimental results on different image data sets, both the effectiveness of hash function and the efficiency of training time of the proposed algorithm have better performance compared with the existing algorithms.

Keywords Deep learning, Image retrieval, Image hashing, Neural network, Optimization algorithm

近年来,随着移动互联网和社交网络媒体等技术的飞速发展,互联网中每天产生数以亿计的图片、视频等多媒体数

到稿日期:2015-06-28 返修日期:2015-09-11 本文受国家自然科学基金(61033010,61272065,61370021,61472453,U1401256),广东省自然科学基金(S2011020001182,S2012010009311,S2013010011905),广东省科技计划项目(2011B040200007,2012A010701013),火烈鸟网络(广州)股份有限公司-中山大学广东省大数据分析与管理重点实验室校企产学研合作研究经费资助。

刘 冶(1989-),男,硕士,主要研究方向为大规模机器学习、社交网络挖掘和深度学习,E-mail:jourkliu@163.com;潘 炎(1979-),男,博士,副教授,硕士生导师,主要研究方向为机器学习、数据挖掘和信息检索;夏榕楷(1988-),男,硕士,主要研究方向为低秩矩阵学习和深度学习;刘 荻(1990-),男,硕士生,主要研究方向为社交网络挖掘和深度学习;印 鉴(1968-),男,博士,教授,博士生导师,主要研究方向为大数据和大数据挖掘,E-mail:issjyin@mail.sysu.edu.cn(通信作者)。

据。在移动互联网时代,人们可以随时随地拍摄各种各样的图片和视频,并在网络上与朋友分享,由此带来的是网络上数字图片和视频的爆炸式增长。目前,社交网络互联网平台 Facebook 每天上传的图片有数亿张,微博和微信等移动互联网应用中每天上传的图片和短视频也数以亿计。相似图片和视频的检索技术是工业界许多实际应用系统中的核心技术,相关的研究受到广泛的关注^[1]。

图像检索技术的核心在于根据用户提供的信息从图像数据库中找出符合用户需求的图片,这项技术的研究从 20 世纪 70 年代就已经开始,并且至今一直都是计算机领域的一个研究热点。在图片搜索引擎(Google、百度、Bing、Yandex)中,通过图片检索技术在海量的互联网图片中检索出与查询图片相似的图片。电子商务网站的应用中,需要在大量商品中找出满足要求的商品,通过手机拍照等方式输入查询图片,以实现购物导航,通过以图搜图的购物搜索技术在多个主要应用(Amazon 购物搜索、百度识图、淘淘搜)中被广泛应用。

在图片和视频检索领域,在海量数据中实现快速、准确的检索是最为核心的研究问题。在各种面向大规模数据的检索方法中,保持相似性哈希(Similarity-preserving Hashing)是近几年兴起的一类代表性方法^[2-4]。以基于保持相似性哈希的图片检索为例,这类方法首先要将每张图片转换成某种特征表示,例如用一个实数特征向量来表示一张图片,再把该特征向量转换成二进制哈希码,最后基于这些哈希码构建哈希表并进行快速检索。

基于图片内容的图像检索技术的关键在于图像特征提取和查找匹配,其中图像特征提取是基础也是关键。图像特征提取包含了对颜色、纹理、形状等图像信息的描述和量化。当特征提取结束后,每一幅图片可以用一个特征向量来表示,后续的检索中判定两张图像的相似性则可以根据图片特征向量之间的相似性来判定。常用的相似性度量有欧氏距离、绝对值距离和马氏距离等,图像检索时通常采用最近邻检索算法(Nearest Neighbor Search)。目前主流的搜索引擎公司(Google、百度)都已经提供基于内容的图像搜索产品(Google Similar Images、百度识图)来满足用户的需求。

基于图片内容的图像检索更加生动且贴近用户的实际需求。由于图像所包含的信息要多于文本信息,基于图片内容的检索技术速度会比基于文本描述的检索技术要慢,基于内容的图像检索技术对硬件的要求也比较高。为了提高检索效率,需要尽可能减少查询时间,基于图片内容的检索还需要与哈希技术相结合。

基于哈希算法的图像检索技术由于可以提高检索效率并降低对硬件的需求而成为研究热点。目前主流的哈希算法将传统的基于内容的图像检索系统中提取到的高维特征通过哈希函数生成一个固定长度的二进制编码(binary code)。采用哈希方法后,一幅图片将由一个固定长度的二进制编码来表示,相似度计算度量可以直接使用海明距离(Hamming Distance),计算速度大大提高,从而加快了系统的检索效率^[7]。

对于基于哈希算法的图像检索技术而言,哈希函数的选择必须尽可能地保证哈希前与哈希后的特征之间的相似度一致,即对于原本相似图像的特征向量,经过哈希函数计算得到的哈希码要尽可能地相似,而不相似图像的特征向量哈希码要尽可能不一样,因而哈希函数的选择至关重要。目前的哈

希算法可以分成两种类型:数据无关(Data-independent)和数据相关(Data-dependent)的哈希算法。数据无关的哈希方法采用的是随机投影(Random Projection)理论,最大的特点是具有通用性。数据相关的哈希方法也称为哈希函数学习(Learning to Hash),这类办法通过挖掘数据的隐含特性,学习到适合该类数据的哈希函数。从机器学习的角度,目前的哈希函数学习算法可以分为以下 3 类:无监督哈希算法、半监督哈希算法和有监督哈希算法。

虽然主流的图像哈希检索算法能够实现图片与图片间的检索并取得不错的效果,但是基于哈希算法的图像检索也存在不足之处,目前主流的方法中图像特征提取与后续的哈希函数学习是分开的,即先对图像提取特征,再对所提取到的特征进行哈希函数学习。如果图像提取得到的特征并不能很好地反映图像与图像间的相似程度,那么后续的哈希函数的选择无论如何也无法满足要求。因此,如何将特征提取与哈希函数的学习结合在一起,从而提高整个检索系统的准确度十分具有研究与应用意义。同时,近年来基于深度学习的有监督哈希算法在效果上取得了一定提升,但由于神经网络的设计比较复杂,同时训练时间较长,直接影响了基于深度学习的哈希算法的广泛应用。

通过上述的分析与研究,本文提出一种基于深度卷积神经网络的快速图像哈希算法,模型的结构如图 1 所示。与现有的哈希算法相比,本文提出的方法有以下优点:1)该方法无需先对图像提取人工特征再进行哈希函数学习,而是直接通过原始图像信息和相似性信息来进行哈希函数的学习;2)通过运用已经训练的神经网络进行迁移学习,可以缩短训练时间,提高哈希函数训练的效率和效果。

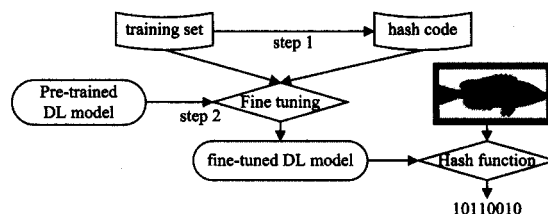


图 1 本文基于深度学习的快速图像哈希算法模型示意图

本文研究工作的主要贡献包括:

- (1)设计了基于深度神经网络的快速图像哈希编码学习框架,提高了哈希编码检索的效果;
- (2)在图像训练集构造哈希编码的求解步骤中设计了快速收敛的优化求解算法;
- (3)选择预训练神经网络在数据集上进行迁移学习的方式,加快了哈希函数训练的速度。

通过多个不同的真实数据集进行实验和结果分析,可以得出所提算法在实验结果上取得了优于基准算法(baseline)的效果和更高的效率,同时提出的哈希算法模型在不同参数设置下具备鲁棒性。

1 相关工作

为了解决海量高维数据下最近邻检索的问题,许多学者提出了近似最近邻检索(Approximate Nearest Neighbor, ANN)^[9,12,14,41],目前已经有很多种数据结构和算法被提出以实现高效的近似最近邻检索。其中一种经典的方法是在 1975 年提出的 KD 树算法^[18],KD 树是一种分割 d 维数据空

间的数据结构,通过对空间的分割,然后在特定空间的部分内进行相关搜索操作从而实现高效的索引。在 KD 树提出之后,陆续有其他的如 M 树^[32]、cover 树^[33]、metric 树^[34]等树形结构被提出来应用到最近邻检索中,但是这些算法也会遇到维数灾难(Curse of Dimensionality)问题。例如用 d 来表示数据的维度,当维度比较小时,可以利用这些数据结构来进行近似最近邻检索的快速求解,如上述提到的 KD 树等。但随着维度的增长,算法的空间复杂度和时间复杂度都呈指数级别增长。

由于速度快、存储空间小等优点,越来越多的学者对基于哈希的检索算法进行了研究。目前主流的哈希算法均基于二进制编码(binary code),这些哈希算法将传统的基于内容的图像检索系统中提取到的高维特征通过哈希函数生成一个固定长度的二进制编码序列,从而大大地降低了内存存储空间,能够把海量图像数据库存储在内存中。图像检索的过程则需要对查询图片和图像数据库中的图像进行运算,计算出相似度,并根据相似度对图像数据库中的图像排序返回结果。哈希后进行查询时可以利用海明距离(Hamming Distance)来判断图像间的相似度,而求解二值哈希码的海明距离也比较高效,从而加快了检索速度,解决了高维及海量数据下的最近邻检索问题。哈希函数的选择在该方法中至关重要,该哈希函数的输出必须是二值编码,同时还必须尽可能地使哈希后与哈希前特征之间的相似度一致。

哈希算法可以看作是一个数据降维的过程,降维的原则是通过某种方法把数据从高维空间映射到低维空间,同时还要尽可能地保留数据在原有的高维空间中的相似性。数据相关的哈希算法是基于学习的(learning-based),与数据有关的这类方法通过挖掘数据本身的隐含特性,学习到适合该类数据的哈希函数。从机器学习的层面来看,根据利用图像标签信息的程度,目前主流的哈希算法又可以分为以下 3 类:1)无监督哈希算法;2)半监督哈希算法;3)有监督哈希算法。

无监督学习(Unsupervised Learning)是一种机器学习的方式,特点是不需要任何标签信息,无监督学习的一种典型方法是聚类(Clustering)算法。无监督哈希学习继承了上述特点,在图像检索中我们无需提供标签信息或图像间的相似度,只需提供表示图像的特征向量,使得算法自动学习目标哈希函数从而生成哈希码。无监督哈希中最具代表性的是局部敏感哈希算法(Locality Sensitive Hashing, LSH),后续的研究也出现了很多其他的无监督哈希算法,例如 KLSH^[10], SH^[11], ITQ^[4]等。

LSH 算法的雏形是由 Indyk 等人^[41]于 1998 年提出,后续由 Gionis^[8]等人在 1999 年改进并进行了深入的分析证明,在 2002 年,Charikar^[2]则提出了基于随机投影的二值哈希算法。可以说 LSH 算法的发明对后续一系列哈希算法的研究发展产生了深远的影响。LSH 算法可以利用随机投影实现,随机投影依据 Johnson-Lindenstrauss 定理^[42]产生特定分布的投影矩阵,然后根据该投影矩阵对原始的高维数据进行线性变换投影到低维空间,并且使得投影后的低维空间在很大程度上都能维持原始高维空间下的距离信息^[17]。

经典的 LSH 算法是在原始的空间上进行随机映射,其要求原始空间是已知的,这对现有的图像检索以及提高其精准度有所限制。2009 年 Kulis 等人提出了 KLSH^[10]算法(Ker-

nalized Locality-sensitive Hashing)来弥补 LSH 的不足, KLSH 通过在核空间而不是原始空间中随机地构造哈希函数,能把 LSH 算法的核心思想推广到使用任意的核函数(Kernel Function)中。

基于随机投影的 LSH 具有数据无关的通用性,但它通常也需要较长的编码长度以提高准确度。为了提高检索系统的效率,人们提出了基于学习的哈希算法,该类型的算法通常只需较短的哈希码就能达到可观的性能。在这其中,谱哈希(Spectral Hashing, SH)^[11]为无监督学习中一种著名的代表性算法。除了要求在哈希后的海明空间中的距离信息要尽量跟原始空间保持一致外,谱哈希还要求得到的哈希码是平衡的,并且每个哈希码间是不相关的。

2011 年, Gong 等人提出了迭代量化哈希算法(Iterative Quantization, ITQ)^[4], ITQ 可以用于监督算法,也可以用于无监督算法。ITQ 通过最小化将数据投影到二值超立方体(Binary Hypercube)顶点中的量化错误来学习一个更好的二进制哈希码。ITQ 中通过旋转数据使其量化错误最小化的做法最早可在文献[1, 7]中发现,不过文献[7]中的做法并没有包含对数据进行降维,而文献[1]中的做法并不是针对二进制哈希码的。ITQ 作者在文中通过实验比较发现,该方法比 LSH 和 SH 有着更高的准确度。

与无监督哈希算法相比,有监督的哈希算法通常都是数据相关的,即不具备通用性。但有监督哈希算法通常能比无监督的哈希算法学习到一个更为紧凑的编码(Compact Code),并且哈希的效果也比无监督的要好^[19, 27-30]。通常在有监督的哈希算法中,随着编码长度的增加,哈希编码检索的效果会提升,然而编码长度增加到一定程度时,再增加长度对效果并不会有很大的提高,有时甚至会使得图像检索效果变差。

MLH(Minimal Loss Hashing)是由 Norouzi 和 Blei 在 2011 年提出的一款有监督哈希算法^[30],其根据结构化 SVM(Structural SVM)构造一个目标函数用于哈希函数的学习^[31]。由于目标损失函数是非连续的(discontinuous)且是非凸的,使得直接优化变得很难,MLH 采用了跟结构化 SVM 类似的做法。

与基于随机投影的方法构造哈希函数的算法不同, Kulis 等提出 BRE(Binary Reconstructive Embedding)算法^[28]。BRE 算法的主要思想是,原始空间中样本间的距离度量信息要尽可能地与哈希后的样本间的距离度量信息一致,这种方法采用了坐标梯度下降的办法迭代地更新哈希函数直到找到一个局部最优解^[28]。

KSH(Kernel-based Supervised Hashing)是在 2012 年由 Liu 等人^[19]提出的哈希算法。与之前的有监督哈希算法如 MLH 和 BRE 不同, KSH 并不是直接通过最小化哈希码间的海明距离来训练哈希函数的,而是通过最小化哈希码的内积,文中证明了最小化哈希码的内积相当于隐式的最小化海明距离,同时,最小化哈希码的内积也使得问题变得更简单,有更好的解法。此外,为了处理数据间非线性可分的情况, KSH 采用了核(kernel)的形式来构造哈希函数。

根据对无监督和有监督哈希算法的优缺点的分析和研究, Wang 等人在 2010 年提出了半监督哈希算法 SSH(Semi-Supervised Hashing)^[15], SSH 利用了有监督哈希和无监督哈希算法的特点来提高哈希函数的质量。

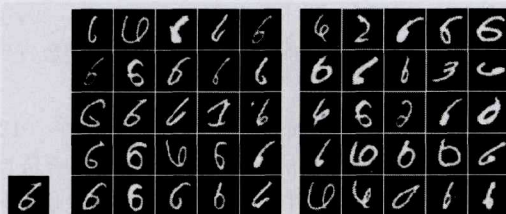
随着大数据技术的发展,近年来深度学习的相关技术也被应用于大规模图像哈希算法中。Xia 等人^[3]设计了基于深度神经网络的两阶段哈希编码学习框架,并在框架提出了具体的哈希算法 CNNH,通过首先构造训练集的哈希编码,再通过深度神经网络训练模型的方式,得到图像检索的哈希函数,Liu 等人具体实现了该哈希算法框架并公开了源代码¹⁾。Lai 等人^[13]针对 CNNH 算法中需要分为两个阶段学习哈希函数的问题,构造了全新的损失函数,提出了基于深度神经网络的单阶段哈希学习框架和算法。基于深度学习的这类算法提高了有监督哈希算法的效果,但也存在一些问题,例如深度神经网络的设计复杂,同时训练时间较长,影响了算法的可扩展性和实用性。

2 图像哈希算法

2.1 算法框架

有监督哈希学习的目标,是从图像训练集学习得到 q 个哈希函数,其中 q 为哈希编码的长度。目前主流的哈希算法基本上都能够实现图片与图片间的检索并取得不错的效果,但其同时也存在着不足之处,这些方法都依赖 GIST^[36]等方法构造人工图像特征,图像特征的提取与后续哈希函数的学习是分开的,即先提取图像特征,然后再对所提取到的特征进行哈希并生成哈希码。如果第一步提取到的特征并不能很好地保留图像与图像间的相似程度,那么后续哈希函数的学习无论如何也无法满足要求。因而,如何将一幅图片编码成为一个有利于哈希函数学习的图像特征也就成为了图像检索中哈希函数学习的重要问题。

通过上述的分析与研究,本文提出一个更适用于图像检索的有监督哈希方法,该方法的输入就是一幅原始图片,最后输出的信息有两个:1)我们的目标哈希函数;2)图像特征提取函数。第一步中输入的相似性矩阵 S 被分解为 H 和 H 的转置,其中 H 表示训练集样本的哈希码,针对矩阵 H 设计了加速的求解优化算法。第二步采用了深层卷积神经网络来学习目标哈希函数以及学习一个能够更加准确地描述图像信息的表示特征,网络架构中选择了已训练好的深度卷积神经网络。输出层的构造可以简单地采用第一步学习到的哈希码作为目标。在查询时,只需将图片输入到网络中,最后通过获取输出层中节点的值即可获得表达该图像的二进制哈希码,图 2 展示了快速图像哈希算法的检索效果。



(a)用于检索的图片 (b) FP-CNNH 的检索结果 (c) KSH 的检索结果

图 2 对比哈希算法在手写数字数据集上的检索效果

2.2 构造训练集哈希编码的优化算法

为便于叙述,先定义哈希算法中的图像相似性矩阵 S :

$$S_{ij} = \begin{cases} +1, & \text{图像 } I_i \text{ 和 } I_j \text{ 在语义上相似} \\ -1, & \text{图像 } I_i \text{ 和 } I_j \text{ 在语义上不相似} \end{cases}$$

对于给定 n 个图像的哈希函数训练集,可以根据图像集合的信息构造相似性矩阵 S 。通过神经网络进行哈希函数的训练且需要图像训练集中的每个图片都有相应的哈希码,因而构造训练集的哈希编码是哈希函数训练的基础,下面介绍提出的构造哈希编码训练集的优化算法的推导求解过程。

首先定义大小为 $n \times q$ 的二值矩阵 H ,并用 H_k 来表示矩阵 H 的第 k 行。我们的目标是设计满足要求的哈希函数,使对于任意给定的两个相似的图像,通过哈希函数得到的哈希编码之间的海明距离要尽可能一致。在文献^[19]中指出哈希码的内积与它们间的海明距离具有一对一的映射关系,因而根据 $H \in \{-1, 1\}^{n \times q}$,可知哈希码内积 $H_i \cdot H_j^T$ 的取值范围为 $[-q, q]$ 。

根据上述定义,可以通过最小化以下目标函数来学习得到图像训练集的哈希编码:

$$\begin{aligned} \min_H \sum_{i=1}^n \sum_{j=1}^n (S_{ij} - \frac{1}{q} H_i \cdot H_j^T)^2 &= \min_H \| S - \frac{1}{q} H H^T \|_F^2 \\ \text{subject to: } H &\in \{-1, 1\}^{n \times q} \end{aligned} \quad (1)$$

其中, $\| \cdot \|_F$ 为弗罗贝尼乌斯范数(Frobenius norm),取值范围为 $H \in \{-1, 1\}^{n \times q}$ 。

对于训练集中的图片,通过对目标函数进行优化求解,得到表示图片的二进制哈希编码的矩阵 H ,同时又使得哈希后它们间的海明距离尽可能与实际的相似程度一致。由于 $H_i \cdot H_j^T$ 的取值范围在 $[-q, q]$ 中,因此将矩阵 $H_i \cdot H_j^T$ 计算结果每个位置上的数值除以 q 使其落在范围 $[-1, 1]$ 之间,使得取值范围与相似性矩阵 S 的取值范围 $S_{ij} \in \{-1, 1\}$ 一致。

虽然定义出了目标函数,但是要直接优化公式依然很困难,由于存在二值约束条件 $H \in \{-1, 1\}^{n \times q}$,使得该问题是 NP 难的,因此与谱哈希(SH)^[11]求解过程类似,需要放松约束,这里将约束条件替换为范围约束条件 $H \in [-1, 1]^{n \times q}$,即将矩阵 H 的取值范围转换为 -1 到 1 之间,而不是原有定义中只有 -1 或 1 的离散取值,那么目标函数就可以转化为以下形式:

$$\begin{aligned} \min_H \| S - \frac{1}{q} H H^T \|_F^2 \\ \text{subject to: } H &\in [-1, 1]^{n \times q} \end{aligned} \quad (2)$$

式(2)依然不容易优化,由于存在项 $H H^T$,使得该问题是非凸的(non-convex)。到目前为止,非凸问题通常采用贪心的策略。我们将采用基于牛顿方向的坐标梯度下降的办法来优化该问题,该方法每次顺序或随机地选取 H 中的一个元素来更新,同时其他元素则保持不变。首先假设当前要更新的元素为 H_{ij} ,那么保持 H 的其他元素不变。使 $H = [H_{\cdot 1}, H_{\cdot 2}, \dots, H_{\cdot q}]$,其中 $H_{\cdot j}$ 表示为矩阵 H 的第 j 列。式(2)可以重写为以下形式:

$$\begin{aligned} \min_H \| S - \frac{1}{q} H H^T \|_F^2 \\ = \min_H \| H_{\cdot j} H_{\cdot j}^T - (qS - \sum_{c \neq j} H_{\cdot c} H_{\cdot c}^T) \|_F^2 \\ = \min_H \| H_{\cdot j} H_{\cdot j}^T - R \|_F^2 = \min_H \sum_{l=1}^n \sum_{k=1}^q (H_{lj} H_{kj} - R_{lk})^2 \\ \text{subject to: } H_{ij} &\in [-1, 1] \end{aligned} \quad (3)$$

其中,为了便于推导和叙述,定义矩阵 R 为:

$$R = qS - \sum_{c \neq j} H_{\cdot c} H_{\cdot c}^T \quad (4)$$

¹⁾ <http://ss.sysu.edu.cn/~py/CNNH/cnnh.html>

由于相似性矩阵 S 是对称矩阵, 可以推导得出 R 也是对称矩阵, 通过固定矩阵 H 中除了 H_{ij} 的其他元素, 可以得到关于 H_{ij} 的目标函数为:

$$\begin{aligned} \min_{H_{ij}} g(H_{ij}) &= \sum_{l=1}^n \sum_{k=1}^n (H_{ij} H_{kj} - R_{lk})^2 \\ &= (H_{ij}^2 - R_{ii})^2 + 2 \sum_{k \neq i} (H_{ij} H_{kj} - R_{ik})^2 + \\ &\quad \text{constant} \end{aligned} \quad (5)$$

subject to: $H_{ij} \in [-1, 1]$

其中用到了矩阵 R 是对称矩阵这样一个事实。假设想要更新 H_{ij} 为 $H_{ij} + d$, 那么对应的优化问题为:

$$\begin{aligned} \min_d g(H_{ij} + d) \\ \text{subject to: } -1 \leq H_{ij} + d \leq 1 \end{aligned} \quad (6)$$

为了方便搜索 d , 利用泰勒展开式将 $g(H_{ij} + d)$ 近似为以下二次函数:

$$g(H_{ij} + d) \approx g(H_{ij}) + g'(H_{ij})d + \frac{1}{2} g''(H_{ij})d^2 \quad (7)$$

其中, $g'(H_{ij})$ 和 $g''(H_{ij})$ 分别为函数 g 在点 H_{ij} 中的一阶导数和二阶导数, 即:

$$\begin{aligned} g'(H_{ij}) &= 4 \sum_{k=1}^n (H_{ij} H_{kj} - R_{lk}) H_{kj} \\ g''(H_{ij}) &= 12 H_{ij}^2 - 4 R_{ii} + 4 \sum_{k \neq i} H_{kj}^2 \end{aligned} \quad (8)$$

通过在近似的二次函数式子中对 d 求导, 并设其导数为 0, 可以得到 d 的解为:

$$d = -\frac{g'(H_{ij})}{g''(H_{ij})} = -\frac{4 \sum_{k=1}^n (H_{ij} H_{kj} - R_{lk}) H_{kj}}{12 H_{ij}^2 - 4 R_{ii} + 4 \sum_{k \neq i} H_{kj}^2} \quad (9)$$

与此同时, 除了满足条件 $-1 \leq H_{ij} + d \leq 1$ 之外, 由哈希编码的二值定义, 还必须约束 d 的取值范围, 使得 $H_{ij} + d$ 的结果为 1 或 -1。由以上可得, d 的取值应当为 $-1 - H_{ij}$ 或 $1 - H_{ij}$, 算法中需要首先计算理想情况下 d 的取值, 然后选择 $-1 - H_{ij}$ 和 $1 - H_{ij}$ 这两个值中理想情况下更接近 d 取值的, 这一步的计算优化直接求解得到 H_{ij} 为 1 或者 -1, 与 CNNH 方法相比, 减少了最终结果还需要取整带来的损失, 同时加快了优化算法的计算速度。

通过式(9)得到 d 的解后, 就得到了 H_{ij} 的更新形式, 即 $H_{ij} \leftarrow H_{ij} + d$ 。

从上述推导中可以发现, 每轮更新 H_{ij} 时需要计算出 d , 计算 d 时最耗时的步骤为求解矩阵 R , 时间复杂度为 $O(n^2)$, 对于较大的训练集, n 可能会很大, 这样会使得每次更新 H_{ij} 的代价过高。因此, 下面讨论如何避免显式地计算 R , 使得每轮更新的时间复杂度从 $O(n^2)$ 降为 $O(n)$ 。

定义一个新的矩阵 $L = HH^T - qS$, 则其一阶导数 $g'(H_{ij})$ 和二阶导数 $g''(H_{ij})$ 可以用如下公式计算:

$$\begin{aligned} g'(H_{ij}) &= 4L_{i \cdot} \cdot H_{\cdot j} \\ g''(H_{ij}) &= 4(H_{\cdot j}^T H_{\cdot j} + H_{ij}^2 + L_{ii}) \end{aligned} \quad (10)$$

其中, $L_{i \cdot}$ 为 L 的第 i 行。可以发现, 当给定矩阵 L 时, 能够在 $O(n)$ 的时间内计算出 $g'(H_{ij})$ 和 $g''(H_{ij})$, 因而 d 的计算可以在 $O(n)$ 的时间内完成。与此同时, 当更新 $H_{ij} \leftarrow H_{ij} + d$ 后, 矩阵 L 只有第 i 行和第 i 列发生变化, 所以可以得到矩阵 L 的更新公式为:

$$\begin{aligned} L_{i \cdot} &\leftarrow L_{i \cdot} + dH_{\cdot j} \\ L_{\cdot i} &\leftarrow L_{\cdot i} + dH_{\cdot j} \\ L_{ii} &\leftarrow L_{ii} + d^2 \end{aligned} \quad (11)$$

由式(11)可以发现矩阵 L 的更新也只需 $O(n)$ 的时间复杂度, 所以只需在算法开始时先构造矩阵 $L = HH^T - qS$, 而在后续每一轮 H_{ij} 的更新以及 L 的更新都能够在 $O(n)$ 时间内完成, 于是每轮的更新复杂度为 $O(n)$ 。

算法 1 基于坐标梯度下降的构造训练集哈希码学习算法

输入: 训练集图片相似性矩阵 $S_{ij} \in \{-1, 1\}^{n \times n}$, 目标哈希码长度 q , 最大迭代次数 T , 精度 ϵ

输出: 长度为 q 的训练集的哈希编码构成的矩阵 H , 其中 H 每个位置的取值范围均为 -1 或 1

初始化: 将 H 随机初始化为 $H_{ij} \in \{-1, 1\}^{n \times q}$;

计算矩阵 $L = HH^T - qS$;

$t \leftarrow 0$; $F_{(0)} \leftarrow \|L\|_F^2$; $H_{(0)} \leftarrow H$; $L_{(0)} \leftarrow L$;

定义集合 $\Omega = \{(1, 1), (1, 2), \dots, (i, j), \dots, (n, q)\}$ 为矩阵 H 的下索引;

For $t=1, 2, \dots, T$ do

 随机打乱集合 Ω 中 $n \times q$ 个元素的顺序;

 For 集合 Ω 中每个元素的下索引 (i, j) do

 ①根据式子计算一阶导数 $g'(H_{ij}) = 4L_{i \cdot} \cdot H_{\cdot j}$ 和二阶导数 $g''(H_{ij}) = 4(H_{\cdot j}^T H_{\cdot j} + H_{ij}^2 + L_{ii})$;

 ②根据梯度计算 d , 选择 $-1 - H_{ij}$ 和 $1 - H_{ij}$ 两个值中与梯度计算得到的 d 取值更接近的一个;

 ③更新 $H_{ij} \leftarrow H_{ij} + d$;

 ④更新 $L_{i \cdot} \leftarrow L_{i \cdot} + dH_{\cdot j}$, $L_{\cdot i} \leftarrow L_{\cdot i} + dH_{\cdot j}$, $L_{ii} \leftarrow L_{ii} + d^2$;

 End For

$t \leftarrow t + 1$;

$F_{(t)} \leftarrow \|L\|_F^2$;

 If $F_{(t)} \leq F_{(t-1)}$ Then

$H_{(t)} \leftarrow H$, $L_{(t)} \leftarrow L$;

 Else

$H_{(t)} \leftarrow H_{(t-1)}$, $L_{(t)} \leftarrow L_{(t-1)}$;

 End If

 If 相对误差 $\frac{F_{(t-1)} - F_{(t)}}{F_{(t-1)}} \leq \epsilon$ Then

 跳出循环体;

 End If

End For

算法 1 给出了基于随机坐标梯度下降的训练集哈希码快速求解算法。算法在内循环中的时间复杂度为 $O(n)$, 整个算法总的时间复杂度为 $O(Tqm^2)$, 其中 T 为最大迭代次数, q 为哈希码的长度, 通常总的迭代次数 T 不需要很大, 而 $q \leq 64$, 于是使得该算法能够适用于 n 相对较大的情况下。

2.3 深度卷积神经网络概述

本文中提出的哈希算法采用了卷积神经网络 (Convolutional Neural Network, CNN)^[20] 的架构, 因而在这里首先简单地介绍下 CNN。CNN 属于前馈神经网络 (Feed-forward Neural Network) 的一种, 它是为识别二维图像而特殊设计的一种多层感知器, 该网络结构对平移、缩放、倾斜、旋转或其他形式地变形具有一定程度的不变性。

神经网络是由大量简单的神经元 (neuron) 通过相互连接而形成的复杂网络结构。简单神经元的数学定义为 $f(x) = \Sigma w_i g_i(x)$, 其中 w 为权重, 是后续需要通过学习而得到的变量, f 为激活函数 (activation), 该神经元的输入为 x , 通过激活函数 f 后得到该神经元的输出为 $f(x)$ 。常用的激活函数有 sigmoid 函数、tanh 函数、rectified linear 函数等。一个神经网络通常由多个神经元组成, 所以多个神经元的输出也可以

作为其他神经元的输入,神经网络中的参数学习可以采用反向传播算法(Back-Propagation algorithm, BP)。

神经网络已被应用在很多领域之中并且取得了显著的效果。在图像识别领域,如果将图片的像素直接输入神经网络则有着明显的缺点,图像对平移、缩放、倾斜、旋转等变形非常敏感。由于上述缺点,大量的图像识别研究是基于人工特征提取设计图像分类算法,具体过程是先对图片进行预处理,提取用来表示该图像的特征,例如 GIST^[36], SIFT^[38], BOG^[39] 等特征,然后再将该特征作为机器学习分类算法的输入。根据问题的不同往往需要人工制定不同的特征或者结合多种特征来提高效果,这些人工特征往往不具有通用性,需要大量时间和经验来获取。

CNN 属于 Hubel-wiesel 架构,基于 Hubel 和 Wiesel 在 1962 年对猫的初级视觉皮层的研究^[26]。日本研究学者 Fukushima 基于该概念提出的神经认知机(neocognitron)是第一个在计算机上进行仿真的模型^[23],但该模型并没有使用全局有监督的训练算法来对模型进行训练。后来 LeCun 等人通过简化架构并采用了反向传播算法(Back-propagation Algorithm)的全局有监督算法来训练整个模型^[20,24,25],并使得该模型在数字识别上取得了令人振奋的效果。

CNN 主要通过 3 种结构来实现对输入具有一定程度的平移或其他变形的不变性:1)局部接受域;2)权值共享;3)子采样(Subsampling 或 Pooling)。CNN 中的卷积层就应用了局部接受域和权值共享这两个想法,并将每层共享权值的神经元组成一个二维平面,称该二维平面为特征图(Feature Map),每个特征图内的权值都是共享的,并将检测某种特定形状的视觉特征。文献^[20,24,25]详细地介绍了关于采用反向传播算法对卷积神经网络的学习过程。

2.4 基于深度神经网络的哈希函数学习

在哈希函数的学习方式上,本文提出的方法与经典方法使用图像的人工特征作为输入的有监督哈希学习方法不同,如图 3 所示,本文的哈希算法在第二阶段使用原始图像作为输入,并利用第一阶段所学习到的训练集的哈希码来进行目标哈希函数的学习。

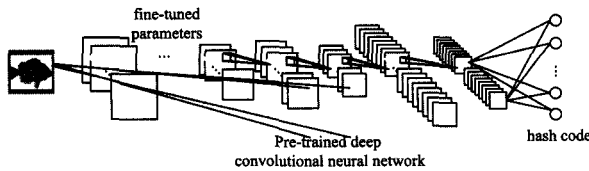


图 3 基于深度卷积神经网络的快速图像哈希训练过程

深度学习(Deep Learning)已经成为机器学习中的一个热门研究领域。在众多的深层结构中,深层卷积神经网络(Deep Convolutional Neural Network, CNN)^[20]是为识别二维图像而特殊设计的一种多层感知器,该网络结构对平移、缩放、倾斜、旋转或其他形式的变形具有一定程度的不变性。CNN 也在许多具有挑战性的视觉任务(如目标检测(object recognition))中展示出了令人振奋的成果^[21]。在第二阶段学习哈希函数的过程中,本文选择了使用深度卷积神经网络的结果来

进行图像哈希函数的学习,将文献^[21]中的神经网络结构作为基础,基于文献^[40]中提出的 Caffe 深度学习框架中的预训练深度卷积神经网络,设计针对哈希函数学习的深度学习算法。同时,在本文设计的深度卷积神经网络中,设置比率为 0.5 的 dropout 技术^[22]也被应用到全连接层。

对于给定一幅图片 I ,得到的 q 个哈希编码值用向量表示为 $b \in [0,1]^q$,在深度神经网络中由于最后一层采用的是 sigmoid 函数,因此输出是在 0 到 1 之间,那么最后用来表示该图片的哈希码为 $sign(b-0.5)$ 。

2.5 哈希算法框架总结

本文提出了一个新的有监督快速图像检索哈希算法,具体算法的流程如下。

(1)通过给定的 n 个已知相似度信息训练图片 $\{I_1, I_2, \dots, I_n\}$,构造相似性矩阵 S ;

(2)将相似性矩阵 S 作为输入,运行提出的基于随机坐标梯度下降的加速优化算法,得到训练图片的二进制哈希码 H ;

(3)利用步骤(2)中学习到的哈希编码矩阵 H 和训练图片集合 $\{I_1, I_2, \dots, I_n\}$ 作为输入预训练的深度卷积神经网络模型来进行哈希函数的学习;

(4)进行图像检索时,将图片作为输入,并获取输出层上红色节点上的值 $b \in [0,1]^q$,最后 $sign(b-0.5)$ 即为该图片的哈希码。

提出的哈希算法无需先对图片提取人工特征,然后再进行哈希学习,而是通过提供的原始图像信息和相似性信息直接学习到最后用来表达这幅图像的哈希函数。同时,除了学习到哈希函数,还学习到一个更能够表示该图像信息的图像表示特征。

3 实验和分析

3.1 实验选取的数据集

为了验证提出的哈希算法的性能,通过真实的数据集对所提算法和主流的算法进行对比实验和分析。

实验一共使用了以下 3 个经典的真实基准数据集:

(1)CIFAR-10¹⁾

该数据集^[35]包含有 60000 张大小为 32×32 的彩色图片。图片分为 10 个类别,包含有猫、汽车、飞机等分类,每个类别有 6000 张图片。

(2)MNIST²⁾

该数据集包含有 70000 张大小为 28×28 的手写数字灰度图片,数字从 0 到 9,每个类别有 7 千张图片。

(3)NUS-WIDE³⁾

该数据集^[16]是一个多标签(multi-label)数据集,一共有 81 个标签(类别),并有接近 270000 张从网络上抓取的图片。每一张图片被赋予一个或多个标签。与文献^[29]一样,这里只考虑在该数据集中出现次数最多的 21 个标签,并只与这 21 个标签有关的图片,每个标签都至少拥有 5000 张图片。

在数据集 CIFAR-10 和 MNIST 中,随机从每一类中抽取

¹⁾ <http://www.cs.toronto.edu/~kriz/cifar.html>

²⁾ <http://yann.lecun.com/exdb/mnist>

³⁾ <http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm>

出 100 张图片作为测试用例,即一共有 1000 张图片作为测试,其余的样本作为实验的图像数据库。对于无监督学习的哈希算法,将排除了测试集后的图像数据库作为训练样本。而对于有监督学习的哈希算法,从每个类别随机抽 500 张作为训练样本,即从图像数据库中抽取了 5000 个样本。由于 CIFAR-10 和 MNIST 数据集都是有标签的,因此可以根据样本的标签来构造相似性矩阵。

在 NUS-WIDE 数据集中,同样随机地从每一类中抽取 100 张图片作为测试用例,由于有 21 个标签信息,因此一共有 2100 张图片作为测试,其余的样本作为图像数据库。与 CIFAR-10 和 MNIST 一致,对于无监督学习的哈希算法,将上述的图像数据库作为训练样本。对于有监督学习的哈希算法,则从图像数据库中对每一类都随机抽取 500 张图片作为训练样本,构成了图片数目为 10500 的训练集。

3.2 对比算法与评估标准

我们将用以下主流哈希算法与所提哈希算法进行比较:1)局部敏感哈希算法(LSH),采用随机投影的哈希算法;2)谱哈希算法(SH);3)无监督迭代量化哈希算法(ITQ);4)最小化损失的有监督哈希算法(MLH);5)二值重构嵌入的有监督哈希算法(BRE);6)有监督的迭代量化哈希算法(ITQ-CCA);7)有监督核哈希算法(KSH);8)基于神经网络的有监督哈希算法(CNNH)。

上述对比的基准算法中,LSH 使用的是我们仔细按照算法流程实现的版本,其他的基准算法都采用作者提供的源码,可以在相关作者的主页上下载。

3.3 实验结果分析

由于提出的算法无需先对图片进行特征提取,因此直接使用图片的像素作为算法的输入。对于其他的对比算法,则根据不同的数据集使用其最常用的特征作为图像表示特征。具体的做法是:对于 CIFAR-10 数据集,用一个 512 维的 GIST 向量^[37]来表示其图像特征^[19]。对于 MNIST 数据集,将图片用一个 784 维的原始图片像素值组成的灰度向量来表示^[29,30]。对于 NUS-WIDE 数据集,则用 500 维的 bag-of-words 特征来表示^[29]。

表 1 在 CIFAR-10 上使用不同长度哈希码的海明排序 MAP

算法	12 位	24 位	32 位	48 位
FP-CNNH	0.612	0.639	0.625	0.616
CNNH	0.438	0.508	0.505	0.519
KSH	0.301	0.331	0.342	0.353
ITQ-CCA	0.261	0.280	0.283	0.287
MLH	0.177	0.191	0.205	0.209
BRE	0.153	0.178	0.191	0.193
SH	0.125	0.126	0.129	0.127
ITQ	0.161	0.165	0.170	0.172
LSH	0.120	0.123	0.119	0.118

实验中使用 MAP(Mean Average Precision)作为结果的评价标准。表 1 显示出 CIFAR-10 数据集上的实验结果,可以发现:在评价标准为 MAP 的情况下,有监督哈希算法大部分的 MAP 都比无监督的效果要好。提出的 FP-CNNH 算法与所有基准算法对比表现出最佳的效果。FP-CNNH 与效果最接近的算法 CNNH 和 KSH 相比,有较大提高,其中 FP-CNNH 算法在哈希编码为 12、24、32、48 位的情况下,分别比基准算法中效果最好的 CNNH 算法提高了 39.73%、25.79%、23.76%、18.69%。其与 KSH 等经典的基准算法对比得到较大幅度的提升,证明了 FP-CNNH 算法通过深度神

经网络框架直接利用图像来学习哈希函数能够更好地保持图像间的相似信息。

在 CIFAR-10 数据集上,将两个阶段的哈希算法模型中的训练时间与 CNNH 进行对比,实验中使用同样的参数设置,分别重复 10 次记录时间并取平均值,可以得出本文设计的第一阶段训练集哈希码快速构造方法,能够比 CNNH 提升 1/10 左右的时间,第二阶段通过预训练的神经网络分类模型针对哈希函数学习目标进行模型训练,达到同样的收敛条件和效果时的哈希函数所花费的时间仅仅是 CNNH 的 1/20 左右。

表 2 示出 MNIST 数据集上的实验结果,其与 CIFAR-10 数据集上的结果类似,可以观察到以下结果:在评价标准为 MAP 的情况下,有监督哈希算法整体上都比无监督的效果要好,并且随着位数的增加,大部分算法的 MAP 也会随之增加。同时,根据所有基准算法的效果可以得出,FP-CNNH 算法的效果最佳,再次证明了 FP-CNNH 直接利用图像来学习哈希函数能够更好地保持图像间的相似信息。

表 2 在 MNIST 上使用不同长度哈希码的海明排序 MAP

算法	12 位	24 位	32 位	48 位
FP-CNNH	0.962	0.965	0.968	0.969
CNNH	0.951	0.955	0.956	0.959
KSH	0.869	0.889	0.891	0.896
ITQ-CCA	0.653	0.686	0.712	0.721
MLH	0.471	0.652	0.650	0.651
BRE	0.512	0.581	0.601	0.622
SH	0.261	0.265	0.258	0.248
ITQ	0.386	0.432	0.421	0.426
LSH	0.185	0.206	0.233	0.241

表 3 是在 NUS-WIDE 数据集上的实验结果,与 MNIST 和 CIFAR 的数据集不同,NUS-WIDE 的图像都是高分辨率图像,比较贴近实际应用,FP-CNNH 依然表现得比其他的基准算法更好。由于 NUS-WIDE 数据集规模较大,为了节省评估 MAP 的计算步骤的时间,采用了检索返回最相似的前 5000 张图片代替整个检索返回的图片集,类似的方法在大规模数据的检索系统评估中常被使用,本文中记作 MP@top5k。可以看到,FP-CNNH 与最接近的算法 CNNH 和 KSH 相比都有一定的提高,证明了提出的基于图像表示学习的想法确实能提高哈希函数的质量,提出的 FP-CNNH 算法比其他的竞争算法有不同程度的提升。

表 3 在 NUS-WIDE 上使用不同长度哈希码的海明排序 MP@top5k

算法	12 位	24 位	32 位	48 位
FP-CNNH	0.622	0.628	0.631	0.625
CNNH	0.618	0.621	0.619	0.620
KSH	0.609	0.615	0.617	0.605
ITQ-CCA	0.552	0.570	0.580	0.582
MLH	0.432	0.436	0.431	0.429
BRE	0.498	0.511	0.518	0.519
SH	0.481	0.520	0.522	0.529
ITQ	0.431	0.423	0.425	0.421
LSH	0.402	0.419	0.423	0.437

结束语 在大规模的图像检索系统中,通常采用近似最近邻的图像检索方法,哈希算法速度快、存储空间小等优点,使其成为海量图像检索中的一项关键图像编码技术。针对目前主要的有监督哈希算法存在的不足,本文提出了一种基于深度卷积神经网络的快速图像哈希算法。该方法通过设计高效的构造图像训练集哈希编码的优化算法,以及选择预训练

的深度卷积神经网络模型训练哈希函数,使得提出的图像检索哈希算法在效率和效果上都取得了相对于基准算法更好的实验结果。

参考文献

- [1] Jégou H, Douze M, Schmid C, et al. Aggregating local descriptors into a compact image representation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010;3304-3311
- [2] Charikar M S. Similarity estimation techniques from rounding algorithms[C]// Proceedings of the Symposium on Theory of Computing. 2002;380-388
- [3] Xia R, Pan Y, Lai H, et al. Supervised Hashing for Image Retrieval via Image Representation Learning[C]// Proceedings of AAAI Conference on Artificial Intelligence (AAAI). 2014
- [4] Gong Y, Lazebnik S. Iterative quantization: A procrustean approach to learning binary codes[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011;817-824
- [5] Wang Hui-feng, Sun Zheng-xing, Wang Jian. Semantic Image Retrieval; Review And Research[J]. Journal of Computer Research and Development, 2002, 39(5):513-523(in Chinese)
王惠锋, 孙正兴, 王箭. 语义图像检索研究进展[J]. 计算机研究与发展, 2002, 39(5):513-523
- [6] Wu Hong, Lu Han-qing, Ma Song-de. A Survey of Relevance Feedback Techniques in Content-Based Image Retrieval[J]. Chinese Journal of Computers, 2005, 28(12):1969-1979(in Chinese)
吴洪, 卢汉清, 马颂德. 基于内容图像检索中相关反馈技术的回顾[J]. 计算机学报, 2005, 28(12):1969-1979
- [7] Jegou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search[C]// Proceedings of the European Conference on Computer Vision. 2008;304-317
- [8] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing[C]// Proceedings of the International Conference on Very Large Data Bases. 1999, 99:518-529
- [9] Kleinberg J M. Two algorithms for nearest-neighbor search in high dimensions[C]// Proceedings of the Symposium on Theory of Computing. New York, ACM Press, 1997;599-608
- [10] Kulis B, Grauman K. Kernelized locality-sensitive hashing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(6):1092-1104
- [11] Weiss Y, Torralba A, Fergus R. Spectral Hashing [J]. Proceedings of the Advances in Neural Information Processing Systems. 2008, 28(3):1753-1756
- [12] Arya S, Mount D M, Netanyahu N S, et al. An optimal algorithm for approximate nearest neighbor searching fixed dimensions [J]. Journal of the ACM, 1998, 45(6):891-923
- [13] Lai H, Pan Y, Liu Y, et al. Simultaneous Feature Learning and Hash Coding with Deep Neural Networks[C]// Proceedings of IEEE International Conference on Pattern Recognition and Computer Vision (CVPR). 2015
- [14] Clarkson K L. An algorithm for approximate closest-point queries[C]// Proceedings of the Symposium on Computational Geometry. 1994;160-164
- [15] Wang J, Kumar S, Chang S F. Semi-supervised hashing for scalable image retrieval[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010;3424-3431
- [16] Chua T S, Tang J, Hong R, et al. NUS-WIDE: a real-world web image database from National University of Singapore[C]// Proceedings of the ACM International Conference on Image and Video Retrieval. 2009;1-9
- [17] Bingham E, Mannila H. Random projection in dimensionality reduction; applications to Image and text data[C]// Proceedings of the International Conference on Knowledge Discovery and Data Mining. 2001;245-250
- [18] Bentley J L. Multidimensional binary search trees used for associative searching[J]. Commun ACM, 1975, 18(9):509-517
- [19] Liu W, Wang J, Ji R, et al. Supervised hashing with kernels [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012;2074-2081
- [20] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324
- [21] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Proceedings of Advances in Neural Information Processing Systems. 2012;1106-1114
- [22] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint, arXiv:1207.0580, 2012
- [23] Fukushima K, Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. Pattern recognition, 1982, 15(6):455-469
- [24] LeCun B B, Denker J S, Henderson D, et al. Handwritten digit recognition with a back-propagation network[C]// Proceedings of Advances in Neural Information Processing Systems. 1990
- [25] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4):541-551
- [26] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1):106-154
- [27] Lin G, Shen C, Suter D, et al. A general two-step approach to learning-based hashing[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013
- [28] Kulis B, Darrell T. Learning to Hash with Binary Reconstructive Embeddings[C]// Proceedings of the Advances in Neural Information Processing Systems. 2009, 22:1042-1050
- [29] Liu W, Wang J, Kumar S, et al. Hashing with graphs[C]// Proceedings of the International Conference on Machine Learning. 2011;1-8
- [30] Norouzi M, Blei D M. Minimal loss hashing for compact binary codes[C]// Proceedings of the International Conference on Machine Learning. 2011;353-360
- [31] Yu C N J, Joachims T. Learning structural svms with latent variables[C]// Proceedings of the International Conference on Machine Learning. 2009;1169-1176
- [32] Ciaccia P, Patella M, Zezula P. M-tree: An Efficient Access Method for Similarity Search in Metric Spaces[C]// Proceedings of the International Conference on Very Large Data Bases. 1997
- [33] Beygelzimer A, Kakade S, Langford J. Cover trees for nearest neighbor[C]// Proceedings of the International Conference on Machine Learning. 2006;97-104

(下转第 51 页)

表3 $\lambda=10^{-3}$ 时算法在数据集上遍历5次的结果

Data Sets	S ² GD+F1			S ² GD+F2			S ² GD+F3			SGD			SGA		
	Time	Fun.	Err.	Time	Fun.	Err.	Time	Fun.	Err.	Time	Fun.	Err.	Time	Fun.	Err.
a9a	1.287	0.426	0.193	1.282	0.4	0.181	1.274	0.389	0.175	0.69	0.507	0.236	0.881	0.449	0.225
gisette	9.501	7.506	0.229	9.374	3.974	0.147	9.331	2.928	0.115	4.523	0.47	0.16	5.961	0.649	0.228

结束语 本文给出了一个基于随机谱梯度的在线学习框架,从优化的角度设计了3种学习算法,其结构简单且易于实施,并通过实验验证了算法的有效性。从算法的设计角度看,所提方法可用于非光滑优化,如基于1范数的学习问题。

参考文献

- [1] Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer, 2000
- [2] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012
- [3] Bottou L. Stochastic Learning[M]// Advanced Lectures on Machine Learning. Springer Berlin Heidelberg, 2004: 146-168
- [4] Bottou L. Large-scale machine learning with stochastic gradient descent[C]// Proceedings of Computational Statistics 2010. Physica-Verlag HD, 2010: 177-186
- [5] Zhang Tong. Solving large scale linear prediction problems using stochastic gradient descent algorithms[C]// Proceedings of the 21st International Conference on Machine Learning. 2004: 919-926
- [6] Robbins H, Monro S. A stochastic approximation method[J]. The Annals of Mathematical Statistics, 1951, 22(3): 400-407
- [7] Powell W B. Approximate Dynamic Programming: Solving the Curses of Dimensionality[M]. John Wiley & Sons, 2007
- [8] Johnson R, Zhang Tong. Accelerating stochastic gradient descent using predictive variance reduction[C]// Advances in Neural Information Processing Systems. 2013: 315-323
- [9] Konecny J, Richtarik P. Semi-stochastic gradient descent methods[OL]. <http://arxiv.org/pdf/1312.1666v2.pdf>
- [10] Le Roux N, Schmidt M, Bach F. A stochastic gradient method with an exponential convergence rate for finite training [C]// Advances in Neural Information Processing Systems. 2012: 2663-2671
- [11] Blatt D, Hero A O, Gauchman H. A convergent incremental gradient method with a constant step size [J]. SIAM Journal on Optimization, 2007, 18(1): 29-51
- [12] Bordes A, Bottou L, Gallinari P. SGD-QN: careful quasi-Newton stochastic gradient descent[J]. Journal of Machine Learning Research, 2009, 10: 1737-1754
- [13] Mokhtari A, Ribeiro A. A dual stochastic DFP algorithm for optimal resource allocation in wireless systems[C]// IEEE 14th Workshop on Signal Processing Advances in Wireless Communications. 2013: 21-25
- [14] Mokhtari A, Ribeiro A. Regularized stochastic BFGS algorithm [C]// IEEE Global Conference on Signal and Information Processing. 2013: 1109-1112
- [15] Schraudolph N, Yu Jin, Günter S. A stochastic quasi-newton method for online convex optimization [C]// International Conference on Artificial Intelligence and Statistics. 2007: 436-443
- [16] Sopyła K, Drozda P. Stochastic gradient descent with Barzilai-Borwein update step for SVM purposes[J]. Information Sciences, 2015, 316(20): 218-233
- [17] Sun Wen-yu, Yuan Ya-xiang. Optimization Theory and Methods; Nonlinear Programming[M]. Springer Science & Business Media, 2006
- [18] Barzilai J, Borwein J M. Two-point step size gradient methods [J]. IMA Journal of Numerical Analysis, 1988, 8(1): 141-148
- [19] Biglari F, Solimanpur M. Scaling on the spectral gradient method [J]. Journal of Optimization Theory and Applications, 2013, 158(2): 626-635
- [20] Farid M, Leong W J, Hassan M A. A new two-step gradient-type method for large-scale unconstrained optimization[J]. Computers and Mathematics with Applications, 2010, 59(10): 3301-3307
- [21] Wright S J, Nowak R D, Figueiredo M A T. Sparse reconstruction by separable approximation[J]. IEEE Transactions on Signal Processing, 2009, 57(7): 2479-2493
- [22] Yu Gao-hang, Guan Lu-tai, Chen Wu-fan. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization[J]. Optimization Methods and Software, 2008, 23(2): 275-293
- [23] Leung W J, Hassan M A. A new two-step gradient-type method for large-scale unconstrained optimization[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [34] Uhlmann J K. Satisfying general proximity / similarity queries with metric trees[J]. Information Processing Letters, 1991, 40(4): 175-179
- [35] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[D]. University of Toronto, 2009
- [36] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145-175
- [37] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3360-3367
- [38] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [39] Csurka G, Dance C, Fan L, et al. Visual categorization with bags of keypoints[J]. Proceedings of the European Conference on Computer Vision, 2004, 1(1-22): 1-2
- [40] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[OL]. arXiv preprint, arXiv: 1408.5093, 2014
- [41] Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality[C]// Proceedings of the Symposium on Theory of Computing. 1998: 604-613
- [42] Johnson W, Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space[J]. Contemporary Mathematics, 1984, 26: 189-206

(上接第46页)

- [34] Uhlmann J K. Satisfying general proximity / similarity queries with metric trees[J]. Information Processing Letters, 1991, 40(4): 175-179
- [35] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[D]. University of Toronto, 2009
- [36] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145-175
- [37] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3360-3367
- [38] Lowe D G. Distinctive image features from scale-invariant key-