

深度学习哈希综述

沈琳林 劫江育娥

(福建师范大学 数学与信息学院 福州 350117)

E-mail: linjie891@163.com

摘要: 深度学习哈希技术将哈希方法与深度学习技术相结合,获得原始数据的特征表示以及语义信息,进而得到数据的二进制哈希码。深度学习哈希方法被广泛应用于图像检索、文本检索以及视频检索等领域,并取得了瞩目成果。本文首先对深度学习哈希方法的研究进展以及方法的基本框架进行了介绍;然后详细阐述了深度学习哈希方法的多种目标函数项,包括重构误差、位平衡、位独立、哈希约束、参数正则化和稀疏性限制等;并根据样本标签的个数(无标签/单标签/多标签)以及输入形式(单样本/二元组/多元组)对该方法进行分类;最后总结了现有深度学习哈希方法的不足以及未来的发展方向。

关键词: 近似最近邻搜索;深度学习哈希;深度学习;哈希方法;目标函数

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2020)10-2082-10

Survey of Deep Learning Hashing

SHEN Lin LIN Jie JIANG Yu-e

(College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, China)

Abstract: Deep learning hashing technology combines the hash method with deep learning technology to obtain the feature representation and semantic information from the original data, and further to obtain the binary hash code of the data. Deep learning hashing method is widely used in image retrieval, text retrieval, video retrieval and other fields with remarkable results. This paper first introduces the state-of-the-art research progress and the basic framework of the deep learning hashing methods. Then it explains in detail a variety of objective functional terms of the deep learning hashing methods, including reconstruction error, bit balance, bit independence, hashing constraint, parameter regularization and sparsity limitation. And it also classifies the deep learning hashing methods based on the number of labels (unlabelled / single label / multi-label) of data and input forms (single sample / duplet samples / multi-tuple samples). Finally, it summarizes the shortcomings of the existing deep learning hashing method and forecasts the prospects for the future research.

Key words: approximate nearest neighbor search (ANN); deep learning hashing (DLH); deep learning (DL); hash method; objective function

1 引言

随着科学技术的发展,世界范围内兴起了“大数据”浪潮,各个领域的数据呈现出井喷式增长的趋势,隐藏在数据之下的信息具有重大价值^[1],因此,有效地对海量数据进行搜索、分析就变得日趋关键。其中最近邻搜索(Nearest Neighbor, NN)^[2-4],即搜索出与查询项距离最近、最相似的数据项,已经成为一项重要的研究内容。随着数据量不断增大,最近邻搜索需要的存储空间越来越大,其计算复杂度越来越高,同时搜索效率也越来越低。因此近似最近邻搜索(Approximate Nearest Neighbor, ANN)^[3-5]由于其准确、高效的性能而得到关注,其中的一项重要技术——哈希方法^[6-8],在大数据中具有搜索效率高、存储成本低、搜索效果准确等优点,渐渐成为研究的一个重要方向。哈希方法的主要思想是将数据点从原始空间映射到汉明空间,在此过程中需保持样本间在原始空间中的相似性,同时数据点被映射为紧凑的二进制编码,从而

降低存储成本,提高搜索速率。

近年来,随着计算机技术的发展,深度学习(Deep Learning, DL)^[7,10,11]迅速成为各学者趋之若鹜的研究热点,被广泛应用于自然语言处理、计算机视觉、机器学习多个领域。深度学习又称深度神经网络,其主要原理是通过利用人工神经网络(Artificial Neural Networks, ANNs)^[10,12,13],对海量数据进行逐层筛选和提取,获得特征表示。

由于哈希算法需要先将数据从原始空间映射到特征空间,即对数据进行特征提取,再将特征映射为哈希码,而深度学习又具有强大的特征学习能力,因此哈希算法逐渐开始利用深度学习进行特征提取,于是衍生出一种新的方法——即深度学习哈希^[7,14-16]。深度学习哈希融合了哈希方法与深度学习技术,通过利用深度学习的方法,获得原始数据的特征表示与语义信息,进而得到数据的二进制哈希码。

目前已有文献对哈希方法展开综述,如Wang^[8]等人对学习哈希的框架和各种类型的代表技术进行了全面的阐述。

收稿日期: 2020-05-29 收修改稿日期: 2020-06-10 基金项目: 国家自然科学基金项目(61472082)资助; 福建省自然科学基金项目(2014J01220)资助。 作者简介: 沈琳,女,1996年生,硕士研究生,研究方向为数据挖掘;林劫,男,1972年生,博士,教授,CCF会员,研究方向为生物信息学和机器学习;江育娥(通讯作者),女,1970年生,博士,副教授,CCF会员,研究方向为数据挖掘、机器学习、生物信息。

文献[9]中对哈希算法的学习进行了全面的综述,并根据保留相似性的方式将其进行分类。而本文关注哈希在深度学习中的应用,介绍了深度学习哈希当前的研究进展并归纳总结了各种不同方法。

2 深度学习哈希介绍

深度学习的快速发展,促使其与哈希技术的结合——深度学习哈希得到越来越多的关注。深度学习哈希的应用越来越广,尤其是在图像检索方面。Salakhutdinov R 等^[17]第一个提出将深度学习与哈希结合,采用无监督的学习方法,利用多个 RBM 提取样本特征,产生哈希码。卷积神经网络哈希^[18] (Convolutional Neural Networks Hashing, CNNH) 方法基于监督的学习方式,将哈希学习过程分解为两个部分:首先学习近似的哈希码,然后基于 CNN 同时学习样本特征和哈希函数。然而,分两步学习哈希码可能会导致得到的是次优解,于是 Lai 等^[19]提出了深度神经网络哈希 (Deep Neural Networks Hashing, DNNH) 方法,该方法基于三元组的有监督学习方法,利用卷积神经网络,同时学习特征和哈希码。为了进一步利用样本的语义监督信息,深度语义排序哈希^[20] (Deep Semantic Ranking Hashing, DSRH) 利用三元组排序损失,以保持图像的语义结构。大多数深度学习哈希方法都是为处理简单的二进制相似性而设计的,与多个标签相关的图像的多级语义结构尚未得到很好的探索。DSRH^[21]通过采用对多级相似性信息进行编码的排序列表保持多标签图像之间的多级语义相似性,从而解决了上述问题。随着数据量的不断增大,为了实现快速图像检索,二进制哈希码深度学习^[22] (Deep Learning of Binary Hash Codes, DLBHC) 在神经网络的倒数第二层加入隐含层,来表示类标签的潜在概念。该方法适用于大型数据集。

随着研究的发展,研究人员不再局限于单一模态的情况,开始对多模态的深度学习哈希方法展开研究。比如基于排序的深度跨模态哈希算法^[23] (Ranking-based Deep Cross-Modal Hashing, RDCMH),首先利用样本的特征和标签信息获得半监督语义排序列表,然后将语义排序信息集成到深度跨模态哈希中。文献[24]提出了一种基于全局和局部语义保持的跨模态检索深度哈希方法 (GLSP),针对模态间和模态内,分别引入局部和全局语义结构,学习哈希码。

3 深度学习哈希过程介绍

深度学习哈希是指通过不断减小损失来训练神经网络,对输入数据提取出高维特征,然后映射为紧凑的二进制哈希码。现有深度学习哈希方法的输入模态大多为图像、文本或视频,广泛应用于图像检索、文本检索以及视频检索等领域。如图 1 所示,深度学习哈希方法的主要组成部分分为输入、深度神经网络结构、产生哈希码的哈希层和损失。常用的神经网络结构有卷积神经网络 (Convolutional Neural Network, CNN)^[25-27]、多层感知机 (Multi-Layer Perception, MLP)^[28-30]、受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM)^[31-33] 和自编码器 (Auto-Encoder)^[34-36] 等。样本的输入形式通常有单样本、二元组、多元组等,不同的输入形式对应的学习方式

不同,学习方式通常分为有监督、无监督和半监督^[62],有监督是指在有样本标签的前提下,通过标签构建样本之间的相似性,从而对网络进行训练,样本标签有单标签和多标签两种情

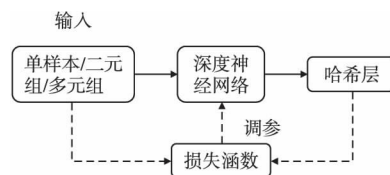


图 1 深度学习哈希方法的主要组成部分

Fig. 1 Main components of deep learning hashing method

况;无监督指在没有样本标签的情况下,利用样本间的空间距离等方式定义样本的空间相似性;半监督通常适用于少量样本有标签,而其余大部分样本无标签的情况。

4 目标函数及误差定义

损失函数用于度量真实值与模型预测值之间的误差,衡量模型正确提取样本特征的能力。通过最小化损失函数对网络结构进行训练,从而得到较优的模型。在深度学习哈希中,特定的任务通常需要制定特定的损失函数,其主要包括以下几个部分:重构误差、哈希码约束、相似性保持、分类误差和松弛化。重构误差用于约束生成的哈希码与原样本之间的误差,即提高模型提取特征的能力;哈希码约束是对生成的哈希码进行约束,使其具有位不相关、位平衡等特性;相似性保持的作用是使样本的相似性信息得到利用,从而生成具有相似性保持的哈希码;分类误差用于提高模型分类样本的准确度,使模型的标签信息进一步得到利用;松弛化的目的在于利用各种方法使模型更易优化。

4.1 重构误差

与自编码器类似的,重构误差指原始数据与经过模型提取的二值哈希码之间的误差。如文献[37]在损失函数中加入重构误差,目的是使网络能够较好地重构原始输入,即精确地提取出原始输入的特征。重构误差常出现在无监督的深度学习哈希方法中,该方法的网络结构通常为自编码器或者生成对抗网络 (Generative Adversarial Networks, GAN),其定义^[38]如公式(1):

$$\text{reconstruct}_{\text{loss}} = \| \mathbf{X} - \mathbf{W}\mathbf{B} - \mathbf{c} \|_{1 \times N}^2 \quad (1)$$

$\mathbf{X} \in \mathbb{R}^{D \times N}$ 表示 N 个 D 维样本的样本矩阵, $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$, \mathbf{W} 是连接网络倒数两层的权重矩阵, \mathbf{A} 表示 \mathbf{X} 的网络输出值矩阵, \mathbf{B} 为 \mathbf{A} 量化后的二值哈希码矩阵, $\mathbf{B} = \text{sgn}(\mathbf{A}) \in \{-1, 1\}^{L \times N}$, \mathbf{c} 指网络最后一层的偏置向量, $\mathbf{1}_{a \times b}$ 是一个大小为 $a \times b$ 且全为 1 的矩阵, $\| \cdot \|_F^2$ 指 Frobenius 范数,激活函数采用符号函数 $\text{sgn}(\cdot)$ 。

4.2 哈希码约束

哈希码约束指在哈希码的层面上进行约束,使生成的哈希码更符合需求,而与样本的输入形式、样本标签的有无,以及所采用的深度神经网络类型均无关。哈希码约束大致包括量化误差、位独立、位平衡、哈希约束、参数正则化以及稀疏性限制。量化误差用于减小特征与量化后的哈希码之间的误差;位独立的作用是使哈希码位与位之间相互独立;位平衡使每

一位哈希码取值 -1 或 1 的概率相等; 哈希约束的目的是约束量化前的实值特征趋近于 +1 或 -1; 参数正则化用于防止模型过拟合; 稀疏性限制在牺牲哈希码长度的前提下, 使生成的哈希码尽可能多地包含零元素。

4.2.1 量化误差

量化误差指量化后的二值码与实值特征之间的差异, Zhang 等人^[39]在损失函数中加入量化误差, 其作用是最小化网络提取的样本特征经量化操作后产生的误差。定义^[23, 38, 40-42, 63, 64]如公式(2)所示:

$$\text{quantization}_{\text{loss}} = \| \mathbf{A} - \mathbf{B} \|_F^2 \quad (2)$$

4.2.2 位独立

为了使生成的哈希码不同的位互不相关, 引入了位独立项^[9]。公式见(3), 其中 \mathbf{I} 是单位矩阵:

$$\text{uncorrelated}_{\text{loss}} = \| \mathbf{B}\mathbf{B}^T - \mathbf{I} \|_F^2 \quad (3)$$

4.2.3 位平衡

为了使生成的哈希码位平衡, 即每一位都有 50% 的概率为 -1 或 1, 提出了位平衡项, 常见的方法有以下几种:

1) 哈希码位内方差最大化

a) 如 Jiang^[41]等人定义的位平衡项:

$$\text{balance}_{\text{loss1}} = \| \mathbf{B}\mathbf{I} \|_2^2 \quad (4)$$

b) 如公式(5)中定义的位平衡项^[65], 其中 $\text{tr}(\cdot)$ 表示矩阵的秩:

$$\text{balance}_{\text{loss2}} = \text{tr}(\mathbf{A}\mathbf{A}^T) \quad (5)$$

2) Zhao^[21]等人定义的位平衡项鼓励哈希码的每个位平均为零, 并确保学习过程更稳定地收敛。公式(6)中 \mathbf{B}_{ij} 为样本 x_j 的第 i 位哈希码, L 指哈希码位数:

$$\text{balance}_{\text{loss3}} = \sum_{i=1}^L \| \text{mean}(\mathbf{B}_{ij}) \|_2^2 \quad (6)$$

4.2.4 哈希约束

哈希约束用于约束网络每个输出节点的输出值, 即准哈希码中的每一个哈希位, 都接近 -1 或 1, 使其最终量化后的误差减小。常见的哈希约束项有以下几种定义, 式中 \mathbf{A}_{ij} 表示样本 x_j 经过网络后得到的第 i 位输出, $\mathbf{1}$ 为元素全为 1 的向量, $\| \cdot \|_1$ 为 L_1 范数, $|\cdot|$ 为对向量每一维取绝对值操作, $\mathbf{A}_{\cdot j}$ 表示样本 x_j 的网络输出值。

1) 如[66]所定义的哈希约束项:

$$\text{const}_{\text{loss1}} = \sum_{i=1}^N \| |\mathbf{A}_{\cdot j}| - 1 \|_1 \quad (7)$$

2) 如[63]定义的哈希约束项:

$$\text{const}_{\text{loss2}} = \min(\| \mathbf{A}_{\cdot j} + 1 \|^2, \| \mathbf{A}_{\cdot j} - 1 \|^2) \quad (8)$$

3) 如 Guo^[43]等人定义的哈希约束项, 其中 $\text{abs}(\cdot)$ 为绝对值函数:

$$\text{const}_{\text{loss3}} = \sum_{i=1}^L \| 1 - \text{abs}(f(\mathbf{A}_{ij})) \|^2 \quad (9)$$

其中,

$$f(\mathbf{A}_{ij}) = \begin{cases} -1, & \mathbf{A}_{ij} \leq -1 \\ 1, & \mathbf{A}_{ij} \geq 1 \\ \mathbf{A}_{ij}, & \text{其他} \end{cases}$$

4.2.5 参数正则化

为了避免模型过拟合, 常用的方法是正则化, 正则化的思想就是在损失函数中加入刻画模型复杂程度的指标^[67]。模型复杂度一般只由权重 \mathbf{W} 决定, 因此引入参数正则化项来最小化模型参数, 其定义^[38]如公式(10)所示:

$$\text{regularization}_{\text{loss}} = \| \mathbf{W} \|_F \quad (10)$$

4.2.6 稀疏性限制

稀疏性限制在牺牲哈希码长度的前提下, 使网络产生的哈希码中含有较少的非零元素, 其目的是从长哈希码中获得高精度, 并且能取得与紧凑哈希码相当的高召回率:

1) 当神经元被激活时, 其输出值接近于 1, 相反, 当被抑制时, 输出近似 0。而稀疏性限制的目的是令神经元在大多数时间内处于抑制状态。定义输入为 x 时, 网络隐含层神经元 i 的激活度为 $a_i(x)$, 则可得出神经元 i 在包含 N 个样本的训练集上的平均激活度 $\hat{\rho}_i$ 为:

$$\hat{\rho}_i = \frac{1}{N} \sum_{i=1}^N a_i(x)$$

令 $\hat{\rho}_i = \rho$, 为了达到稀疏性限制的目的, 通常使稀疏性参数 ρ 接近 0。如公式(11)引入基于相对熵的因子^[68]:

$$\text{sparse}_{\text{loss1}} = \sum_{i=1}^s \text{KL}(\rho \| \hat{\rho}_i) = \sum_{i=1}^s \rho \log \frac{\rho}{\hat{\rho}_i} + (1 - \rho) \frac{1 - \rho}{1 - \hat{\rho}_i} \quad (11)$$

其中 s 表示隐含层神经元的个数, $\text{KL}(\rho \| \hat{\rho}_i)$ 代表分别以 ρ 和 $\hat{\rho}_i$ 为均值的两个伯努利随机变量之间的相对熵。

2) 公式(12)^[44]中利用 L_1 范数正则化项实现稀疏性限制, 其中 α 为控制稀疏性程度的参数, $\mathbf{B}_{\cdot i}$ 是样本 x_i 的哈希码:

$$\text{sparse}_{\text{loss2}} = \alpha \| \mathbf{B}_{\cdot i} \|_1 \quad (12)$$

4.3 相似性保持

相似性保持指在原始空间中相似的样本, 其在汉明空间中的二值哈希码之间的汉明距离较小, 而不相似的样本, 二值码汉明距离较大。样本间的相似性度量可根据样本之间距离的远近或者标签是否相同来定义。相似性保持通常应用在有监督的深度哈希方法中, 而在无监督的情况下, 需要样本间的空间距离作为监督信息。相似性保持项通常分为样本的空间相似性和样本语义相似性。

4.3.1 样本空间相似性

样本空间相似性通常是在样本无标签的情况下, 利用样本空间的距离, 度量样本相似性, 进而保持哈希码的相似性。大致有以下两种方法:

1) 利用拉普拉斯矩阵保持相似性, 如公式(13)定义的相似性保持项^[38]:

$$\text{sp}_{\text{loss1}} = \text{tr}(\mathbf{A}\mathbf{L}\mathbf{A}^T) \quad (13)$$

其中拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{R}$ 保存了关于初始空间中样本间相似性的信息; 对角矩阵 \mathbf{D} , 其对角元素值为矩阵 \mathbf{R} 行或列 $d_{jj} = \sum_i r_{ij}$, 权重矩阵 $\mathbf{R} \in \mathbb{R}^{N \times N}$, 其元素 r_{ij} 表示样本 x_i 与 x_j 之间的距离:

$$r_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right)$$

2) 基于图模型构造样本间的相似度矩阵 \mathbf{S} , 将此矩阵作为样本的相似性信息, 用于指导二值哈希码的生成, 使其能够保持样本的相似性。其做法是根据原始空间中样本的特征, 找到与其最有可能相似或标签相同的样本。如文献[65]多次使用 K 最近邻算法 (k -Nearest Neighbor, k NN), 构建出样本的最近邻图谱, 然后在此图谱上进行搜索, 扩展图谱上所有节点的最近邻。具体做法为: 首先计算两两样本之间的余弦距离,

用 KNN 搜索保留每个样本的前 K_1 个相似样本, 得到初始的相似矩阵, 用 S_1 表示:

$$(S_1)_{ij} = \begin{cases} 1 & x_j \text{ 是 } x_i \text{ 的 } K_1 \text{ NN} \\ 0 & \text{其他} \end{cases}$$

然后在 S_1 的基础上再进行一次 KNN, 此时是计算样本 K_2 个最近邻之间的相似度, 得到 S_2 :

$$(S_2)_{ij} = \begin{cases} 1 & x_j \text{ 是 } x_i \text{ 的 } K_2 \text{ NN} \\ 0 & \text{其他} \end{cases}$$

最后将 S_1 与 S_2 合并, 得到最近邻图谱 S :

$$S_{ij} = \begin{cases} 1 & (S_1)_{ij} = 1 \text{ 或 } x_j \text{ 是 } x_i \text{ 的 } K_2 \text{ NN 的 } K_1 \text{ NN} \\ -1 & \text{其他} \end{cases}$$

4.3.2 语义相似性

1) 利用样本间哈希码相似性与标签相似性之间的差异定义语义相似性: 对于 $L \times N$ 的哈希码矩阵 B , 相似性保持项如下^[39]:

$$sp_{loss2} = \|LS - BB^T\|_F^2 \quad (14)$$

其中 $S = \{s_{ij}\}_{N \times N}$ 是样本的相似性矩阵:

$$\begin{cases} s_{ij} = 1 & x_i \text{ 和 } x_j \text{ 相似} \\ s_{ij} = 0 & \text{其他} \end{cases}$$

2) 通过最小化交叉熵损失函数保持相似性: 如文献[41]定义的交叉熵损失函数. 样本 x_i 与 x_j 的哈希码 $B_{\cdot i}$ 和 $B_{\cdot j}$ 之间的相似性 $\Omega_{ij} = \frac{1}{2} B_{\cdot i}^T B_{\cdot j}$, Ω_{ij} 越大, 则表示两个哈希码越相似. 使用 $\text{sigmoid}(\cdot)$ 对 Ω_{ij} 进行阈值化. 样本对的哈希码与相似度的似然为:

$$p(s_{ij} | B) = \begin{cases} \sigma(\Omega_{ij}), & s_{ij} = 1 \\ 1 - \sigma(\Omega_{ij}), & s_{ij} = 0 \end{cases}$$

若哈希码越相似, 即 $\sigma(\Omega_{ij})$ 越大, 则似然 $p(s_{ij} | B)$ 越大, 反之越小. 对似然取负对数即为交叉熵损失函数:

$$sp_{loss3} = - \sum_{s_{ij} \in S} \log p(s_{ij} | B) = - \sum_{s_{ij} \in S} [s_{ij} \Omega_{ij} - \log(1 + \exp(\Omega_{ij}))] \quad (15)$$

3) 通过令相似样本哈希码之间的汉明距离最小化, 同时使不相似样本哈希码之间的距离最大化, 保持语义相似性:

a) 为了最小化类内方差, 最大化类间方差, 定义如下相似性保持项^[40]:

$$sp_{loss4} = \text{tr}(\sum_B - \sum_W) \quad (16)$$

其中, $\sum_B = \frac{1}{N_D} \text{tr}((A_{d1} - A_{d2})(A_{d1} - A_{d2})^T)$, $\sum_W =$

$$\frac{1}{N_S} \text{tr}((A_{s1} - A_{s2})(A_{s1} - A_{s2})^T).$$

N_D 和 N_S 分别是不相似样本对数和相似样本对数, $\{A_{d1}, A_{d2}\}$ 表示不相似样本对集合 D 中样本对的网络输出, $\{A_{s1}, A_{s2}\}$ 表示相似样本对集合 S 中样本对的网络输出.

b) 如下所定义的相似性保持项中^[45] 第一项惩罚相似样本对映射到不同的二值码, 保证了相似样本之间的距离尽可能地小. 第二项惩罚不相似的样本对映射到相近的二值码, 使得不相似样本之间的距离尽可能大, 其中 M 是边缘阈值参数, 用于限制样本对之间的距离, 以平衡两项损失, 当 $\|A_{\cdot i} - A_{\cdot j}\|_1 > M$, 认为两个样本不相似, 即不属于同一类.

$$sp_{loss5} = s_{ij} \|B_{\cdot i} - B_{\cdot j}\|_1 + (1 - s_{ij}) \max\{0, M - \|B_{\cdot i} - B_{\cdot j}\|_1\} \quad (17)$$

4) 通过最小化语义空间中的联合概率分布 P 与汉明空间中的联合概率分布 V 之间的 Kullback-Leibler 分歧, 保持样本的语义相似性. 如 Lei^[24] 等人利用样本之间共同标签的数量作为监督信息, 为每个样本学习相似性保持哈希码. 首先定义 G 为具有 N 个样本的数据集的关联矩阵, 其元素 G_{ij} 为样本 x_i 与 x_j 的公共标签个数. 然后利用 G 计算语义空间中的联合概率分布 P , 其元素 p_{ij} :

$$p_{ij} = \frac{G_{ij}}{\sum_{i=1}^N \sum_{j=1, j \neq i}^N G_{ij}} \quad p_{ii} = 0$$

在汉明空间中, 利用汉明距离 $d_H(\cdot, \cdot)$ 计算概率分布 V , 将其元素 v_{ij} 定义为 $B_{\cdot i}$ 和 $B_{\cdot j}$ 的联合概率分布, v_{ij} 越大表明哈希码 $B_{\cdot i}$ 与 $B_{\cdot j}$ 越相似:

$$v_{ij} = \frac{\exp(-d_H(B_{\cdot i}, B_{\cdot j}))}{\sum_{k=1}^N \sum_{m=1, m \neq k}^N \exp(-d_H(B_{\cdot k}, B_{\cdot m}))} \quad v_{ii} = 0$$

于是 KL 分歧为:

$$sp_{loss6} = KL(P \| V) = \sum_{i=1}^N \sum_{j=1, j \neq i}^N p_{ij} \log \frac{p_{ij}}{v_{ij}} \quad (18)$$

5) 通过语义排序的方法, 利用样本之间不同的相似性程度, 使生成的哈希码也具有相应的相似性. 如 Zhao^[21] 等人采用语义排序的方法, 定义 r 为数据集中样本 x 与查询项 q 的语义相似性水平. 当样本与 q 的所有标签都一样的, 则其相似性水平最高, 为所有标签个数. 以此类推, 相似性水平递减, 当样本与 q 无公共标签时, 其相似性水平最低, 为 0. 根据样本与 q 的相似性水平, 对数据集的所有样本进行降序排序, 得到关于 q 的排序列表 $\{x_i\}_{i=1}^l$, l 为列表长度. 然后根据排序列表定义排序损失:

$$sp_{loss7}(B_{\cdot q}, \{B_{\cdot i}\}_{i=1}^l) = \sum_{i=1}^l \sum_{j: r_j < r_i} [\delta d_H(B_{\cdot q}, B_{\cdot i}, B_{\cdot j}) + M]_+ \quad (19)$$

其中, $[\cdot]_+ = \max(0, \cdot)$, $\delta d_H(B_{\cdot q}, B_{\cdot i}, B_{\cdot j}) = d_H(B_{\cdot q}, B_{\cdot i}) - d_H(B_{\cdot q}, B_{\cdot j})$. 参数 M 是边缘阈值, 控制样本对距离的最小边缘.

4.4 分类损失

样本标签不仅用于挖掘图像语义结构的监督信息, 还能提供样本分类信息. 在损失函数中加入分类损失能够减小样本分类错误产生的误差. 分类损失大致有以下两种情况:

1) 利用 softmax 损失来定义样本 x_i 的分类损失. 若模型对于样本真正所属类别的预测概率为 1, 则无损失, 否则将产生 softmax 损失:

$$\text{softmax}_{loss} = - \frac{1}{N} \sum_{i=1}^N \log \frac{\exp(W_y^T x_i)}{\sum_{j=1}^c \exp(W_j^T x_i)} \quad (20)$$

其中 c 表示样本类别数, y 为样本 x_i 的标签, W 表示权重.

2) 利用交叉熵定义样本 x 的分类损失:

$$\text{crossEn} = - \sum_{i=1}^l (y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)) \quad (21)$$

其中 $y_i \in \{0, 1\}$ 是二值标签向量, l 指标签数, \hat{p}_i 表示对于样本 x 关于标签 i 的预测概率.

3) 将全连接层看作分类器, 其作用是把二值码分到对应的类. 如公式(22)所示, 将二值码的学习看作是线性的分类问题, 以最小化分类误差^[46]:

$$\text{classified}_{loss} = \|Y - BW\|^2 \quad (22)$$

4.5 松弛化

在对网络输出进行量化时,通常使用符号函数进行二值化,然而符号函数不可导,使得网络优化困难,因此通常使用以下几种方法进行松弛化:

1) 如公式(23)定义的松弛化项^[47,48,66],用双正切激活函数 $\tanh(\cdot)$ 替代符号函数 $\text{sgn}(\cdot)$,将输出映射到 $[-1, 1]$ 范围内,其中 \mathbf{A}_{ij} 为样本 x_j 的第 i 位网络输出, β 为控制平滑度的参数,其作用是决定函数值接近 -1 或者 1 的速度,可通过逐渐增大 β 来近似 $\text{sgn}(\cdot)$:

$$\tanh(\mathbf{A}_{ij}) = \frac{1 - \exp(-\beta \mathbf{A}_{ij})}{1 + \exp(-\beta \mathbf{A}_{ij})}, i = 1, 2, \dots, L \quad (23)$$

2) 如 4.3.2-b) 所示,样本 x_i 与 x_j 的哈希码 $\mathbf{B}_{\cdot i}$ 和 $\mathbf{B}_{\cdot j}$ 之间的相似性为 $\Omega_{ij} = \frac{1}{2} \mathbf{B}_{\cdot i}^T \mathbf{B}_{\cdot j}$, Li^[42] 将此式子中离散的二值码 \mathbf{B} 替换为实数网络输出 \mathbf{A} , 将 Ω_{ij} 替换为 φ_{ij} ,即用连续值代替离散值,以便优化:

$$\varphi_{ij} = \frac{1}{2} \mathbf{A}_{\cdot i}^T \mathbf{A}_{\cdot j} \quad (24)$$

3) 文献[69]中将哈希码之间的距离由汉明距离替换成欧氏距离,并在损失函数中增加量化误差项,以此替换二值化约束:

4) 除了将汉明距离替换成欧氏距离之外,可以添加如下正则化项替代二值约束来实现松弛化^[49]:

$$\|\mathbf{A}_{\cdot j} - 1\|_1 \quad (25)$$

5 深度学习哈希方法分类

如图 2 所示,深度学习哈希方法可以根据样本的标签情况,分为无标签的深度学习哈希方法、单标签的深度学习哈希方法和多标签的深度学习哈希方法,本文将对每一类方法分为单样本、二元组、多元组进行详细的介绍。

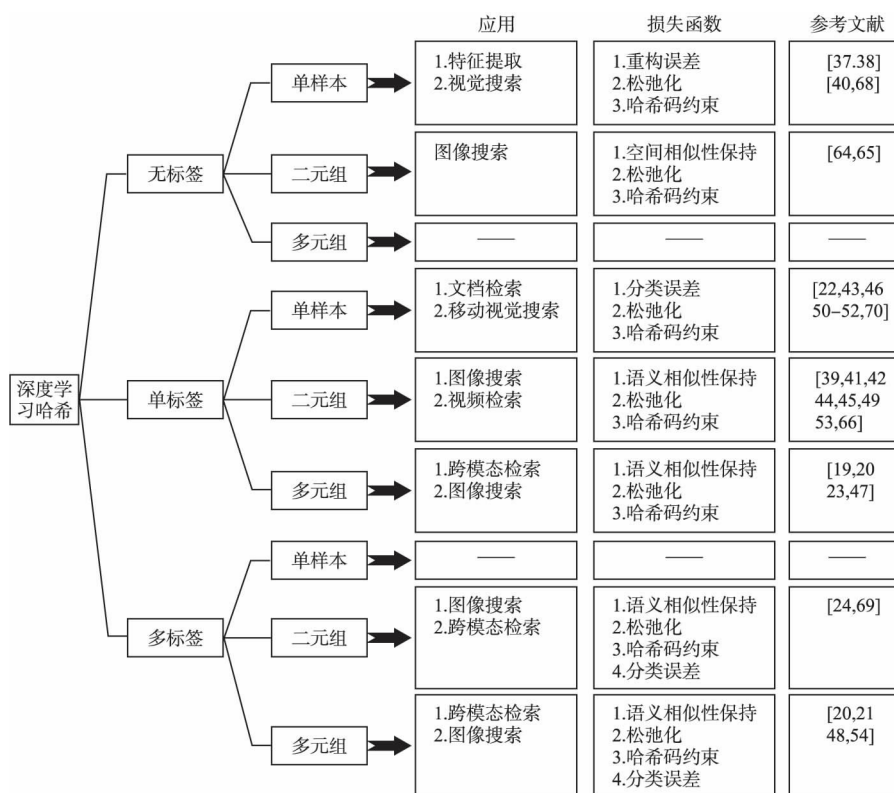


图 2 深度学习哈希方法分类

Fig. 2 Classification of deep learning hash method

5.1 无标签的深度学习哈希方法

无标签的深度学习哈希方法指在没有标签的情况下,利用深度学习进行哈希的一类方法。该方法通常有两类做法,一类是不需要监督信息,因此无法利用样本间的相似性关系,生成的哈希码也没有相似性保持特性;另一类是在没有标签的情况下,利用样本间的空间距离衡量样本的相似性,常用的方法有通过 KNN 等方法生成样本的最近邻图谱,或者构建拉普拉斯矩阵等。

5.1.1 输入样本形式为单样本

单样本输入的无标签深度学习哈希方法指样本的输入形式为一个样本,经过神经网络结构,通过最小化设计的损失函

数训练网络,最后将样本映射为哈希码。大部分单样本输入的无标签深度学习哈希方法^[37,40,68]考虑到输入形式以及无标签的限制,都不考虑相似性保持特性,更多的使用重构误差、哈希码约束来定义损失函数,但部分论文^[38,63]引入了拉普拉斯矩阵来保持样本的空间相似性。此类方法多应用于样本特征的提取或可视化搜索等。

如 Xia^[37] 在异构深度哈希框架下,利用堆栈自编码器 (Stacked Auto Encoder, SAE) 对特征与二值码之间的非线性映射进行建模,即生成初始的二值码,该层的损失函数考虑在位平衡和位独立的约束下,最小化重构误差。然后,利用受限玻尔兹曼机层最大化训练样本的似然,来减小初始二值码的

维数. 最后使用 P-R 曲线验证该方法的有效性.

5.1.2 输入样本形式为二元组

二元组输入的无标签深度学习哈希方法指将成对的样本输入到网络中, 利用样本间的空间相似性关系、哈希码约束等目标函数项来训练网络, 最终生成两个哈希码的一类深度学习哈希方法^[64, 65]. 由于二元组的输入形式以及样本无标签, 该类方法往往会利用样本间的空间相似性关系, 在损失函数中加入相似性保持项, 使生成的二值哈希码具有语义保持性.

如文献[64]首先将 KNN 构建的近邻图谱作为监督信息, 利用 CNN 生成样本对应的哈希码; 然后将其输入到生成对抗网络中, 并输入二进制噪声, 利用生成器、判别器以及损失函数进行训练, 优化哈希码. 总体损失函数为相邻结构损失、内容损失和对抗性损失之和. 其中, 相邻结构损失是 KNN 构造的相似矩阵与生成的哈希码之间的差异, 其作用是使二值码具有相似性保持; 内容损失定义为像素化均方差损失、VGG 损失、量化损失之和.

5.1.3 输入样本形式为多元组

据了解, 目前尚无针对多元组输入的无标签深度学习哈希方法的研究. 由于多元组输入形式一般是为了通过利用样本之间的语义相似性关系, 更好地保持样本的语义信息. 而对于无标签的情况, 可以利用样本之间的空间相似性定义多元组相似性, 并利用哈希码约束等损失项定义损失函数, 从而生成相似性保持的哈希码. 但该方法很难确定样本相似或不相似的阈值, 并且很难界定多个样本之间的相似性等级. 若使用样本的距离来衡量样本的相似性关系, 则需要计算多元组样本两两之间的距离. 因此, 若在今后的研究工作中, 如果能够解决样本距离的计算问题, 则可对多元组输入的无标签深度学习哈希方法进行探索.

5.2 单标签的深度学习哈希方法

单标签的深度学习哈希方法指在样本只有单一标签时, 利用深度学习进行哈希的方法. 通常使用样本标签确定样本的相似性关系, 当样本的标签一样时, 则认为样本相似, 反之, 当样本标签不一致时, 则样本不相似. 单标签的深度学习哈希方法通常通过保持样本的语义来生成二值哈希码.

5.2.1 输入样本形式为单样本

单样本输入的单标签深度学习哈希方法^[22, 43, 46, 50-52, 70]指输入单样本到深度神经网络中, 通过定义的损失函数得到有效的二值哈希码. 该类方法常利用分类误差、哈希码约束作为损失函数项来训练网络, 受其输入形式的限制, 大部分方法不具有相似性保持能力. 单样本输入的单标签深度学习哈希方法常应用于移动视觉搜索、文档搜索等场景.

如 Guo^[43]提出了一种卷积神经网络 (CNN) 与哈希学习相结合的方法, 通过引入一个新的损失层和一个新的哈希层, 学习保存语义信息的哈希函数. 在该方法中, 使用卷积层学习具有语义保持特性的特征, 然后利用 Softmax 损失层从样本的类标签中学习有区分性的特征, 最后通过哈希层和 hinge softmax 损失层学习更类似哈希的编码. hinge softmax 损失如公式 (26) 所示, 其中 P 为每个类的预测概率, y 代表样本的真实标签, \tilde{y} 指代剩下的所有标签, m 是一个惩罚松弛度参数:

$$\text{hingeSoftmax}_{\text{loss}}(y, p) = \begin{cases} 0 & p_y \geq \max(\tilde{p}_{\tilde{y}}) + m \\ -\log(p_y) & p_y < \max(\tilde{p}_{\tilde{y}}) + m \end{cases} \quad (26)$$

5.2.2 输入样本形式为二元组

二元组的单标签深度学习哈希方法^[39, 41, 42, 44, 45, 49, 53, 66]指将成对的样本输入到深度神经网络中, 利用样本间的语义相似性关系训练网络, 进而生成哈希码的一类方法. 由于带单标签的二元组输入形式, 该类方法往往会根据样本标签的异同定义样本的语义相似性关系, 并在损失函数中加入相似性保持项、哈希码约束项, 使生成的二值哈希码具有语义保持性. 此类方法常用于视频检索、图像检索等.

如 Li^[42]提出了深度成对监督哈希 (deep pairwise-supervised hashing, DPSH) 的端到端学习框架, 将成对的图像输入到 CNN 中, 利用基于样本语义相似性的损失函数, 训练 CNN, 产生最终的哈希码. 该框架所设计的损失函数如下: 通过在损失函数中加入量化误差项来最小化量化损失, 同时利用图像的标签定义相似性保持项, 考虑到优化问题, 对损失函数进行松弛化. 最后在两个数据集上测试该方法的 mAP 值.

5.2.3 输入样本形式为多元组

目前的多元组深度学习哈希方法^[19, 20, 23, 47]大多指输入形式为三元组的方法. 三元组指一次输入三个样本, 第一个样本为主样本, 第二个样本与第一个样本相似, 第三个样本与第一个样本不相似. 根据样本相似性关系指导生成对应关系的哈希码, 即第一个样本的哈希码与第二个样本哈希码的距离小于与第三个样本哈希码的距离. 多元组输入的单标签深度学习哈希方法通常利用哈希码约束等损失函数项, 并保持样本的语义关系, 指导深度神经网络生成具有相似性保持的哈希码. 该类方法常被应用在跨模态检索、图像检索上.

如 Lai^[19]等提出了一种基于三元组排序损失的监督哈希方法, 输入三元组图像到 CNN 中, 提取图像特征, 通过最小化三元组排序损失训练网络, 然后利用划分和编码模块将提取的特征分为多个分支, 每个分支对应一个哈希位.

5.3 多标签的深度学习哈希方法

多标签的深度学习哈希方法指的是当样本具有多个标签时, 根据样本的语义排序关系是否有公共标签, 定义样本的语义相似性, 然后通过深度学习对样本进行哈希, 产生语义保持的哈希码. 样本的语义排序关系可以通过对数据库样本与查询样本之间共同标签个数的多少进行排序得到.

5.3.1 输入样本形式为单样本

据了解, 目前尚无针对单样本输入的多标签深度学习哈希方法的相关研究, 可通过 softmax 分类损失以及哈希码约束定义损失函数来保持样本的语义信息, 但由于样本具有多个标签, 需要多个 softmax 分类损失, 较为繁琐, 并且无法保持样本与样本之间的语义相似性, 因此对于多标签的情况, 大部分采用二元组或多元组的输入形式, 不仅能通过 softmax 分类损失保持样本的语义信息, 并且能充分利用样本的语义相似性关系.

5.3.2 输入样本形式为二元组

二元组输入的多标签深度学习哈希方法^[24, 69]通常根据样本是否具有公共标签定义其相似性, 若两个样本之间至少有一个公共标签, 则认为其相似, 否则不相似. 其损失函数项大致有样本的语义相似性保持项、哈希码约束和分类损失等. 此类方法通常用于跨模态检索等场景.

如 Lei^[24]提出了一种基于全局和局部语义保持的跨模态

检索深度哈希方法. 针对图像形式的输入, 利用 CNN 提取特征. 针对文本形式的输入, 利用 MLP 提取文本特征. 在多标签的情况下, 对于不同模态, 利用样本的相似性(是否有共同标签)与所学哈希码的汉明距离之间的关系, 使所学习的哈希码保留局部语义结构; 对于模态内, 引入具有全局多级相似性的监督信息, 最小化语义空间和汉明空间中的联合概率分布之间的 Kullback-Leibler 分歧, 使每个模态的哈希码具有全局语义保持性.

5.3.3 输入样本形式为多元组

多元组输入的多标签深度学习哈希方法^[20 21 48 54]通常是指在样本具有多标签的情况下, 根据数据集中样本与主样本的公共标签个数, 对数据集中所有样本进行排序, 得到主样本的排序列表, 并根据排序顺序定义其与主样本的相似性水平. 大多数多元组输入的多标签深度学习哈希方法都采用三元组的输入形式. 在此三元组中, 第一个样本作为主样本, 第二个样本与第一个样本的相似性水平高于第三个样本与第一个样本的相似性水平, 以此作为语义相似性保持的约束条件, 再加上分类损失、哈希码约束等损失函数项, 训练深度神经网络. 该类方法适用于跨模态检索等应用.

如基于顺序敏感的深度哈希算法^[48], 利用卷积神经网络提取特征, 然后通过一个哈希层产生二值哈希码, 并利用基于三元组的排序损失以及基于交叉熵的多标签分类损失来挖掘多标签样本的信息, 最后通过松弛化减少量化损失. 其基于三元组的排序损失为:

$$L_R(x_q) = \sum_{i=1}^l \sum_{j: r_j < r_i} \frac{2^{r_i} - 2^{r_j}}{Z} \max(0, d_H(b_i, b_j) - d_H(b_i, b_j) + M) \quad (27)$$

其中 x_q 为主样本, l 为 x_q 的排序列表的大小, r_i 是关于样本 x_i 与 x_q 的语义相似性水平, M 控制两对样本对的最小边缘距离, Z 是与排序列表长度有关的常数.

6 评估指标

深度学习哈希方法的评估指标有 mAP、precision、recall、ACG 等, 用于评估方法的性能、准确率. 令 N_r 表示检索到的实际与查询样本相似的样本个数, N_s 表示搜索到的样本总个数, N_a 指的是数据集中所有与查询样本相似的样本数.

6.1 查准率(Precision)

查准率(Precision)是指利用深度学习哈希方法检索到的样本中与查询样本相似的样本占有检测到的样本的比例, Xu^[55]等用此指标来评估模型的检索性能:

$$\text{Precision} = \frac{N_r}{N_s} \quad (28)$$

6.2 查全率(Recall)

查全率(Recall)是指深度学习哈希方法检索出的样本数(即相似于查询样本)占数据集中所有符合条件(即与查询样本相似)的样本数的比例, 用于评估方法检索的全面性:

$$\text{Recall} = \frac{N_r}{N_a} \quad (29)$$

6.3 F 值(F Score)

由于查准率与查全率有互逆性, 即当查准率高时, 查全率

可能处于较低的值, 反之, 于是引入 F 值来综合评估查准率和查全率:

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times (\text{Precision} + \text{Recall})} \quad (30)$$

当 $\beta > 1$ 时, 说明在当前的学习任务中, 较关注查全率, 反之, 当 $\beta < 1$ 时, 比较重视查准率, 当 $\beta = 1$ 时, 即 F1 值, 查准率与查全率同等重要, 需要同时考虑.

6.4 查准率-查全率曲线(P-R Curve)

根据深度学习哈希方法的预测结果对样本进行排序, 依次取不同样本作为划分阈值, 则可以得到多组 Precision 和 Recall, 绘制得到 P-R 曲线. P-R 曲线越靠近坐标(1, 1)越好, 此时查准率和查全率都接近 1. 若方法 A 的 P-R 曲线完全位于另外一个方法 B 的下方, 即被完全包住, 则认为方法 B 优于方法 A. 若两条 P-R 曲线交叉, 则需要根据 P-R 曲线下面积 AP(Average Precision)的大小或者平衡点的位置进行判断, AP 越大, 则方法的搜索性能越好. Torralba^[56]等应用 P-R 曲线, 比较在哈希码长度不同的情况下, 模型性能的差异.

6.5 受试者工作特征曲线(ROC)

ROC 曲线是受试者工作特征曲线(Receiver Operating Characteristic), 与 P-R 曲线相似, 将样本按照深度学习哈希方法学习得到的结果进行排序, 然后依次将每个样本当作正例, 每次得到一组真正例率(True Positive Rate, TPR)和假正例率(False Positive Rate, FPR), 并分别作为纵坐标和横坐标, 最终得到 ROC 曲线.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (31)$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (32)$$

ROC 曲线适用于分类任务, 用于评估方法分类的准确性. 若方法 A 的 ROC 曲线处于方法 B 的范围内, 则认为方法 B 优于方法 A. 若两条 ROC 曲线相交, 则需要根据 ROC 曲线下的面积 AUC(Area Under ROC Curve)大小进行判断, AUC 越大, 则方法的分类效果越好.

6.6 平均精度(AP)

AP(Average Precision)是平均精度, 在检索应用中, 表示深度学习哈希方法对查询样本 q 的返回结果的平均查准率. 其中 Precision_i 表示前 i 个返回的检索结果的查准率, r_i 指第 i 个返回结果与查询样本 q 是否相关, 相关时 $r_i = 1$, 不相关时为 0, n 是检索返回的样本数, N 为数据集中与 q 相关的图片总数.

$$\text{AP}_q = \frac{\sum_{i=1}^k \text{Precision}_i \times r_i}{N} \quad (33)$$

6.7 平均精度均值(mAP)

mAP(mean Average Precision)是平均精度均值, 即所有查询样本 $Q = \{q_1, \dots, q_m\}$ 的 AP 的均值, 用于评估方法的平均检索性能, mAP 越大, 则方法的性能越好. Dong^[57]等用 mAP 来评估模型用于人脸识别^[70]的性能.

$$\text{mAP} = \frac{\sum_{q \in Q} \text{AP}_q}{|Q|} \quad (34)$$

6.8 归一化的累计收益折扣(NDCG)

NDCG(Normalized Discounted Cumulative Gain)通常适

用于语义排序的情况,用于评估样本的排序效果.给定查询样本 q 及其具有 p 个样本的排序列表, NDCG 分数如公式(35)所示,其中 r_i 为排序列表中第 i 个返回样本与查询样本 q 的相似性等级,如公共标签个数等, Z 为标准化因素.

$$\text{NDCG}@p = \frac{1}{Z} \sum_{i=1}^p \frac{2^{r_i} - 1}{\log(i + 1)} \quad (35)$$

6.9 平均累计收益 (ACG)

ACG (Average Cumulative Gain) 通过计算前 p 个位置内样本相似性等级的平均值.

$$\text{ACG}@p = \frac{1}{p} \sum_{i=1}^p r_i \quad (36)$$

7 总结与展望

本文对深度学习哈希方法的研究进展以及方法的基本框架进行了介绍;然后详细阐述了深度学习哈希方法的多种目标函数项,包括重构误差、位平衡、位独立、哈希约束、参数正则化和稀疏性限制等;并根据样本标签的个数(无标签/单标签/多标签)以及输入形式(单样本/二元组/多元组)对该方法进行分类,详细介绍每类方法的主要过程.

综上所述,现有的深度学习哈希方法从输入形式到学习方式,从神经网络结构类型到应用场景都得到了深入的研究.单元组、二元组和多元组的输入形式,有监督、半监督和无监督的学习方式,CNN、RBM 和 GAN 等神经网络结构,单模态和多模态的应用等都使得深度学习哈希方法在图像搜索^[58]、文本搜索^[59]和视频检索^[60]等各个方面得到了广泛的研究与应用.但仍存在以下几个不足,未来的研究工作可以针对这几个方面展开:

1) 深度学习哈希方法由于深度神经网络具有强大的提取特征的能力以及二值哈希码的特性,取得了较好的性能,得到了较为广泛的应用.特别是在图像搜索方面,然而对于其他方面的应用仍然研究较少,未来工作可以针对除图像、视频等常见的领域之外,如生物序列方面、文本方面等进行研究,将深度学习哈希方法推广应用到更广泛的领域.

2) 现有的深度学习哈希方法大多数仅使用 CNN、RBM 提取样本特征,对于其他神经网络的应用有待研究.未来可根据不同的应用需求,使用 LSTM、RNN 等其他神经网络.

3) 进一步地,由于最后通常采用不可导的符号函数 $\text{sgn}(\cdot)$ 来产生二值码,因此存在网络的优化问题,现有的优化方法仍有待改进.今后的工作方向可针对此问题加以研究.

4) 现有的深度学习哈希方法大多数采用有监督的学习方法,在样本无标签或有标签的样本数较少时,所能采用的无监督^[61]以及半监督方法较少.因此,在将来的工作中可对此方面进行研究.

对于多元组输入的深度学习哈希方法,目前大多数仍是三元组的输入形式,将来可对其他多元组(大于三元组)输入形式的方法加以探索,以应用在其他场景,如当样本具有多个标签,且标签具有大小等级关系时,可考虑采用多元组的输入形式.

References:

[1] Mayer-Schönberger V, Cukier K. Big data: a revolution that will

transform how we live, work, and think [M]. Boston: Eamon Dolan/Houghton Mifflin Harcourt, 2013.

- [2] Shakhnarovich G, Darrell T, Indyk P. Nearest-neighbor methods in learning and vision [J]. Pattern Analysis & Applications, 2008, 19(2): 377. doi: 10.1109/TNN.2008.917504.
- [3] Dubey A, Maaten L, Yalniz Z, et al. Defense against adversarial images using web-scale nearest-neighbor search [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 8767-8776.
- [4] Prokhorenkova L. Graph-based nearest neighbor search: from practice to theory [J]. arXiv preprint arXiv: 1907.00845, 2019.
- [5] Liu T, Moore A W, Gray A, et al. An investigation of practical approximate nearest neighbor algorithms [C]//International Conference on Neural Information Processing Systems, MIT Press, 2004.
- [6] Liu W, Wang J, Kumar S, et al. Hashing with graphs [C]//Proceedings of the 28th International Conference on Machine Learning (ICML), Washington, 2011: 1-8.
- [7] Yan C, Xie H, Yang D, et al. Supervised hash coding with deep neural network for environment perception of intelligent vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(1): 284-295.
- [8] Wang J, Liu W, Kumar S, et al. Learning to hash for indexing big data—a survey [J]. Proceedings of the IEEE, 2015, 104(1): 34-57.
- [9] Wang J, Zhang T, Sebe N, et al. A survey on learning to hash [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 769-790.
- [10] Schmidhuber J. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61: 85-117.
- [11] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis [J]. Medical Image Analysis, 2017, 42(9): 60-88.
- [12] Dayhoff J E, Deleo J M. Artificial neural networks [J]. Cancer, 2001, 91(S8): 1615-1635.
- [13] Alanis A Y. Electricity prices forecasting using artificial neural networks [J]. IEEE Latin America Transactions, 2018, 16(1): 105-111.
- [14] Cao Z, Long M, Wang J, et al. Hashnet: deep learning to hash by continuation [C]//IEEE International Conference on Computer Vision, 2017.
- [15] Lu X, Song L, Xie R, et al. Deep hash learning for efficient image retrieval [C]//IEEE International Conference on Multimedia & Expo Workshops, 2017.
- [16] Kang Y, Kim S, Choi S. Deep learning to hash with multiple representations [C]//IEEE International Conference on Data Mining, 2013.
- [17] Salakhutdinov R, Hinton G. Semantic hashing [J]. International Journal of Approximate Reasoning, 2009, 50(7): 969-978.
- [18] Xia R, Pan Y, Lai H, et al. Supervised hashing for image retrieval via image representation learning [C]//Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI Press, 2014: 2156-2162.
- [19] Lai H, Pan Y, Liu Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3270-3278.
- [20] Yao T, Long F, Mei T, et al. Deep semantic-preserving and rank-

- ing-based hashing for image retrieval [C]//International Joint Conference on Artificial Intelligence (IJCAI) 2016: 3931-3937.
- [21] Zhao F ,Huang Y ,Wang L ,et al. Deep semantic ranking based hashing for multi-label image retrieval [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,2015: 1556-1564.
- [22] Lin K ,Yang H F ,Hsiao J H ,et al. Deep learning of binary hash codes for fast image retrieval [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops , 2015: 27-35.
- [23] Liu X ,Yu G ,Domeniconi C ,et al. Ranking-based deep cross-modal hashing [J]. arXiv preprint arXiv: 1905.04450 2019.
- [24] Ma L ,Li H ,Meng F ,et al. Global and local semantics-preserving based deep hashing for cross-modal retrieval [J]. Neuro Computing 2018 312: 49-62.
- [25] Krizhevsky A ,Sutskever I ,Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems 2012: 1097-1105.
- [26] Szegedy C ,Liu W ,Jia Y ,et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015: 1-9.
- [27] Lin M ,Chen Q ,Yan S. Network in network [J]. arXiv preprint arXiv: 1312.4400 2013.
- [28] Ruck D W ,Rogers S K ,Kabricky M. Feature selection using a multilayer perceptron [J]. Journal of Neural Network Computing , 1990 2(2) : 40-48.
- [29] L Liu G Z. Identification method of rice leaf blast using multilayer perception neural network [J]. Transactions of the Chinese Society of Agricultural Engineering 2009 25: 213-217.
- [30] Abd Wahab M H ,Sudin N ,Sulaiman M S ,et al. A model to predict diabetic patient using multi-layer perceptron [J]. Universiti Tun Hussein Onn Malaysia 2010.
- [31] Zhang C X ,Ji N ,Wang G. Restricted boltzmann machines [J]. Chinese Journal of Engineering Mathematics ,2015 ,32(2) : 159-173.
- [32] Salakhutdinov R ,Mnih A ,Hinton G. Restricted boltzmann machines for collaborative filtering [C]//International Conference on Machine Learning 2007.
- [33] Fischer A ,Igel C. An introduction to restricted Boltzmann machines [C]//Iberoamerican Congress on Pattern Recognition ,Springer , Berlin ,Heidelberg 2012: 14-36.
- [34] Hsu C C ,Hwang H T ,Wu Y C ,et al. Voice conversion from non-parallel corpora using variational auto-encoder [C]//Signal & Information Processing Association Summit & Conference 2017.
- [35] Shi X ,Zhu Y ,Ning X ,et al. Transformer fault diagnosis based on deep auto-encoder network [J]. Electric Power Automation Equipment 2016 36(5) : 122-126.
- [36] Zhang L ,Lu Y ,Wang B ,et al. Sparse auto-encoder with smoothed l1 regularization [J]. Neural Processing Letters 2018 47(3) : 829-839.
- [37] Xia Z ,Feng X ,Peng J ,et al. Unsupervised deep hashing for large-scale visual search [C]//Sixth International Conference on Image Processing Theory ,Tools and Applications (IPTA) ,IEEE ,2016: 1-5.
- [38] Bai C ,Wang W ,Zhao T ,et al. Deep learning compact binary codes for fingerprint indexing [J]. Frontiers of Information Technology & Electronic Engineering 2018 ,19(9) : 1112-1123.
- [39] Zhang S ,Li J ,Jiang M ,et al. Scalable discrete supervised multimedia hash learning with clustering [J]. IEEE Transactions on Circuits and Systems for Video Technology 2017 28(10) : 2716-2729.
- [40] Erin Liong V ,Lu J ,Wang G ,et al. Deep hashing for compact binary codes learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015: 2475-2483.
- [41] Jiang Q Y ,Li W J. Deep cross-modal hashing [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2017: 3232-3240.
- [42] Li W J ,Wang S ,Kang W C. Feature learning based deep supervised hashing with pairwise labels [J]. arXiv preprint arXiv: 1511.03855 2015.
- [43] Guo J ,Zhang S ,Li J. Hash learning with convolutional neural networks for semantic based image retrieval [C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining ,Springer , Cham 2016: 227-238.
- [44] Masci J ,Bronstein A M ,Bronstein M M ,et al. Sparse similarity-preserving hashing [J]. arXiv preprint arXiv: 1312.5479 2013.
- [45] Zhong D ,Shao H ,Liu Y. Hand dorsal vein recognition based on deep hash network [C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV) ,Springer ,Cham 2018: 26-37.
- [46] Zhou X ,Shen F ,Liu L ,et al. Graph convolutional network hashing [J]. IEEE Transactions on Cybernetics 2018 49(9) : 1-13.
- [47] 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.
- [48] Chen Z ,Cai R ,Lu J ,et al. Order-sensitive deep hashing for multimorbidity medical image retrieval [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention , Springer ,Cham 2018: 620-628.
- [49] Liu H ,Wang R ,Shan S ,et al. Deep supervised hashing for fast image retrieval [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016: 2064-2072.
- [50] Yang H F ,Lin K ,Chen C S. Supervised learning of semantics-preserving hashing via deep neural networks for large-scale image search [J]. arXiv preprint arXiv: 1507.00101 2015 ,1(2) : 3 ,doi: 10.1109/TPAMI.2017.2666812.
- [51] Li Y ,Wan L ,Fu T ,et al. Piecewise supervised deep hashing for image retrieval [J]. Multimedia Tools and Applications ,2019 ,78(17) : 1-21.
- [52] Yang H F ,Tu C H ,Chen C S. Adaptive labeling for deep learning to hash [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2019.
- [53] Jin L ,Shu X ,Li K ,et al. Deep ordinal hashing with spatial attention [J]. IEEE Transactions on Image Processing ,2018 ,28(5) : 2173-2186.
- [54] Bepler T ,Berger B. Learning protein sequence embeddings using information from structure [J]. arXiv preprint arXiv: 1902.08661 , 2019.
- [55] Xu Jia-ming ,Wang Peng ,Tian Guan-hua ,et al. Convolutional neural networks for text hashing [C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI2015) 2015.

- [56] Torralba A ,Fergus R ,Weiss Y. Small codes and large image data-bases for recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition ,Anchorage ,AK 2008: 1-8.
- [57] Dong Z ,Jing C ,Pei M ,et al. Deep CNN based binary hash video representations for face retrieval[J]. Pattern Recognition 2018 81: 357-369.
- [58] Duan Y ,Lu J ,Wang Z ,et al. Learning deep binary descriptor with multi-quantization [C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2017.
- [59] Shen Y ,He X ,Gao J ,et al. Learning semantic representations using convolutional neural networks for web search [C]//Proceedings of the 23rd International Conference on World Wide Web ,ACM , 2014: 373-374.
- [60] Liong V E ,Lu J ,Tan Y P ,et al. Deep video hashing [J]. IEEE Transactions on Multimedia 2017 19(6) : 1209-1219.
- [61] Shen F ,Xu Y ,Liu L ,et al. Unsupervised deep hashing with similarity-adaptive and discrete optimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018 40(9) : 1-1.
- [62] Li Wu-jun ,Zhou Zhi-hua. Learning to hash for big data: current status and future trends [J]. Chinese Science Bull 2015 60(5-6) : 485-490.
- [63] Chen Zhao-jia. Triplet-based deep hashing network for cross-modal retrieval [D]. Xi'an: Xidian University 2018.
- [64] Peng Yan-fei ,Wu Hong ,Zi Ling-ling. Image retrieval based on hash method and generative adversarial networks [J]. Laser & Optoelectronics Progress 2018 55(10) : 98-104.
- [65] He Tao. Research on image retrieval methods based on hashing and deep learning [D]. Chengdu: University of Electronic Science and Technology of China 2018.
- [66] Wang Hai-long ,Yu Jing ,Xiao Chuang-bo. Deep non-relaxation hashing based on point pair similarity [J/OL]. ACTA AUTOMATICA Sinica 2019 <http://kns.cnki.net/kcms/detail/11.2109.TP.20190402.1428.004.html>.
- [67] Zhen Ze-yu ,Gu Si-yu. Tensorflow practice Google deep learning framework [M]. Beijing: Publishing House of Electronics Industry , 2017.
- [68] Ou Xin-yu ,Wu Jia ,Zhu Heng ,et al. Image hashing retrieval method based on deep self-learning [J]. Computer Engineering & Science 2015 37(12) : 2386-2392.
- [69] Li Tong. Deep hashing method based multi-label image retrieval [D]. Beijing: Beijing University of Posts and Telecommunications 2018.
- [70] Zhang Dong-ping ,Yin Yi-bo. Pedestrian reidentification on deep hash algorithm [J]. Journal of China University of Metrology , 2017 28(2) : 208-213.

附中文参考文献:

- [62] 李武军 ,周志华. 大数据哈希学习: 现状与趋势 [J]. 科学通报 , 2015 60(5-6) : 485-490.
- [63] 陈兆佳. 基于三元组深度哈希的跨模态检索方法 [D]. 西安: 西安电子科技大学 2018.
- [64] 彭晏飞 ,武 宏 ,董玲玲. 基于哈希算法及生成对抗网络的图像检索 [J]. 激光与光电子学进展 2018 55(10) : 98-104.
- [65] 何 涛. 基于深度学习和哈希的图像检索的方法研究 [D]. 成都: 电子科技大学 2018.
- [66] 汪海龙 ,禹 晶 ,肖创柏. 基于点对相似度的深度非松弛哈希算法 [J/OL]. 自动化学报 2019 <http://kns.cnki.net/kcms/detail/11.2109.TP.20190402.1428.004.html>.
- [67] 郑泽宇 ,顾思宇. TensorFlow 实战 Google 深度学习框架 [M]. 北京: 电子工业出版社 2017.
- [68] 欧新宇 ,伍 嘉 ,朱 恒 ,等. 基于深度自学习的图像哈希检索方法 [J]. 计算机工程与科学 2015 37(12) : 2386-2392.
- [69] 李 通. 基于深度哈希的多标签图像检索算法研究 [D]. 北京: 北京邮电大学 2018.
- [70] 章东平 ,尹奕博. 深度哈希算法行人再识别技术研究 [J]. 中国计量大学学报 2017 28(2) : 208-213.