

深度跨模态哈希综述

钱 伟

重庆三峡学院 重庆万州 404020

摘要: 深度跨模态哈希检索是基于深度学习的多模态检索技术,旨在用一种模态的数据去检索另一种模态的内容相似的数据。例如,通过图像检索文本或者以图像检索视频等。跨模态哈希检索的关键任务是:不同模态数据特征向量提取,不同模态数据的特征向量的相似性度量。

关键字: 深度学习; 跨模态检索; 哈希检索

A review of deep cross-modal hashing

Wei Qian

Chongqing Three Gorges University Chongqing Wanzhou 404020

Abstract: Deep cross-modal hashing retrieval is a multi-modal retrieval technology based on deep learning, which aims to use data from one modality to retrieve data with similar content from another modality. For example, text retrieval by image or video retrieval by image etc. The key tasks of cross-modal hashing retrieval are: feature vector extraction of different modal data, and similarity measurement of feature vectors of different modal data.

Keywords: Deep learning ; Cross-modal hashing ; Hashing retrieval

前言

随着大数据时代的来临,网络上不同模态的多媒体数据量激增,人们难以准确且高效地获取相关信息,在这样的背景下,跨模态哈希检索一直是学术界研究的热点。本文从多模态哈希检索技术、标准数据集、评价指标以及跨模态哈希检索技术未来的发展趋势四个方面叙述跨模态哈希的研究现状。

跨模态哈希检索问题定义

数学定义	含义
$X = \{x_1, \dots, x_n\} \in \mathbb{R}^{d_x \times n}$	图像样本集
$Y = \{y_1, \dots, y_n\} \in \mathbb{R}^{d_y \times n}$	文本样本集
$L = \{l_1, \dots, l_n\} \in \mathbb{R}^{d_l \times n}$	样本标签
$f_x(x_i, \theta_x)$	图像特征向量
$f_y(y_i, \theta_y)$	文本特征向量
$B_x = \text{sign}(f_x(x_i, \theta_x))$	图像样本哈希码
$B_y = \text{sign}(f_y(y_i, \theta_y))$	文本样本哈希码
$y_q = \text{minsim}(x_q, y_t)$	相似性度量函数

表 1 跨模态哈希检索问题的符号定义

跨模态哈希检索的目标是学习不同模态的数据的特征并用二值表示,图像和文本数据的特征被提取之后,通过哈希函数(sign)转换成二进制码投影到公共的汉明空间(Hamming Space)中进行相似性度量。二进制码表示的特征向量能够节省大量存储空间,且能够提供高效的检索速度,因此受到了广泛关注。

一、深度跨模态哈希检索的主要方法

跨模态哈希检索旨在学习两个特征提取函数,用于提取两个模态的特征信息。基于深度学习的跨模态哈希检索具有高效性以及节省存储成本受到了学术界的广泛关注。

1. 有监督方法

有监督方法旨在利用人工标注的语义标签信息对学习过程进行监督,从而提高算法的性能。

Jiang 等人提出深度跨模态哈希(deep cross-modal hashing, DCMH)将特征学习和哈希码学习集成到同一框架中,在这个端到端的框架中,每个模态对应一个神经网络,同时学习特征表示,同时由语义标签得到相似性矩阵(similarity matrix, S),利用相似性矩阵监督特征学习过程,若两个样本相似则,反之为 0。该方法取得了较好的性能,但是每个模态的数据是由单独的网络进行学习,并未考虑之间的联系。

Li 等人提出自监督对抗哈希方法(Self-Supervised Adversarial Hashing Networks for Cross-Modal Retrieval, SSAH),使用两个对抗网络联合学习高位特征以及在不同模态下的对应哈希码。一方面使用对抗学习来监督最大化不同模态之间语义关联和特征不一致性;另一方面无缝添加一个自监督的语义网络,来发现多标签标注中的语义信息。

Zou 等人提出基于多标签增强的自监督跨模态哈希算法(Multi-label enhancement based self-supervised deep cross-modal hashing, MESDCH),该方法首先提出了一个多标签语义亲和力和保持模块,该模块使用 Relu 变换

来统一学习哈希码表示的相似性和原始实例对应的多标签语义相似性,并定义了一个正约束损失函数来保持其相似性,从而进一步提高了跨模态哈希检索的性能。

2. 无监督方法

虽然有监督方法能够取得较好的性能,但是考虑到数据集的样本量日益剧增,对样本语义进行标注通常需要耗费大量的资源,因此导致的成本增高。无监督方法因为不需要语义标签的监督,因此更加适用于真实场景中,且推动了实际应用的发展。

Zhang 等人提出了无监督对抗跨模态哈希方法(Unsupervised Generative Adversarial Cross-modal Hashing,UGACH),利用对抗生成网络(Generative Adversarial Network,GAN)对不同模态的数据进行特征学习。

Yu 等人提出了图近邻相关保持方法(Deep Graph-neighbor Coherence Preserving Network for Unsupervised Cross-modal Hashing,DGCPN)该方法衍生自图模型,通过考虑数据与近邻的图近邻关系,主要考虑了图近邻关联,一致性关联以及模态内和模态间关联,进而探索数据在途中的固定性质,进而提高准确性。

Zhu 等人提出了线性跨模态哈希检索方法提高检索效率(Linear Cross-Modal Hashing for Efficient Multimedia Search,LCMH),考虑到模态内与模态间的相似性,将训练数据划分为K个簇,然后用数据到K个簇的质心的距离来表示数据点。

跨模态哈希检索旨在不断改进算法模型,进而寻找最优算法。跨模态哈希检索将数据从高维空间向低维空间投影,不可避免地导致了信息的丢失,且在特征由实值向二值转换时,也会造成一定的信息缺失。随着数据量的增大,算法模型会难以避免地产生过拟合的现象。这些问题是该领域亟待突破的重点。

二、跨模态哈希检索常用数据集和评估指标

数据集对于跨模态哈希检索性能的影响十分重要,该领域常用数据集主要有以下几种:

1.NUS-WIDE:

NUS-WIDE 数据集由 269468 个实例组成,其中包含 81 个独特类别的图像和对应的文本。在实验中,通常使用一个 1000 维的词包向量来表示每个文本数据。数据集中,每个图像实例约有 6 个标签与之对应。该数据集的数据量较大,但是局限于仅有两种模态的数据。

2.MIRFLICKR-25 K

该数据集中共有 25000 个图像文本对,每个实例都属于所提供的 24 个类中的一个。通常会排除少于 20 个文本标记的实例,实验中使用的样本数量约为 20000 个实例。

3.MICROSOFT COCO2014

该数据集由两个子集组成,一个是包含 82785 个用于训练的图像的集合,另外一个包含 40504 个实例的测试集,每个图像有 5 个标签,并使用 80 个类别中的至少一个进行注释。

4.IAPRTC-12

该数据集包括 2000 个图像文本对,每一对都用 275 个类中的至少一个进行注释。

跨模态哈希检索最常用的两种检索方式是汉明排序和哈希查找。汉明排序方式根据汉明距离给出给定查询的排名列表,而哈希查找方式从查询点检索特定汉明半径内的实例。有两种常用的衡量标准(平均精度(mean average precision(MAP)),topN 精度曲线(topN Curves))。其中 MAP 的计算方式如下:

其中表示第 n 个样本的检索精度,则检索结果与样本相关,反之为 0。

三、跨模态哈希检索发展趋势展望

跨模态哈希检索属于热门领域,离实际应用的要求还有较大距离,主要有以下几个问题需要解决:

1. 考虑多标签信息的充分利用,在有监督学习中,每个样本共享的标签不同,需要设计合理的方法去判别样本之间的相似性。

2. 弥合语义鸿沟,在特征提取阶段,不同的网络结构会忽略原始语义的部分信息,在后续的相似性计算过程中,不能够得到最准确的表示。

3. 建立样本之间更深层次的联系,如文本数据需要考虑前后词语的时间联系,图片样本中前景与背景信息的权重。

参考文献:

[1] 侯腾达,金冉,王晏祯,蒋义凯.跨模态检索研究综述[J/OL].计算机工程与应用:1-14[2022-10-16].<http://kns.cnki.net/kcms/detail/11.2127.tp.20220817.1612.002.html>

[2] 徐文婉,周小平,王佳.跨模态检索技术研究综述[J/OL].[1]计算机工程与应用:1-14[2022-10-16].<http://kns.cnki.net/kcms/detail/11.2127.TP.20220818.1451.008.html>

[3] 欧卫华,刘彬,周永辉,宣瑞晟.跨模态检索研究综述[J].贵州师范大学学报(自然科学版),2018,36(02):114-120.DOI:10.16614/j.cnki.issn1004-5570.2018.02.019.

[4] 尹奇跃,黄岩,张俊格,吴书,王亮.基于深度学习的跨模态检索综述[J].中国图象图形学报,2021,26(06):1368-1388.