



图像多标签学习的研究概述

摘要

随着图像大数据的爆发,特别是用户贡献数据的飞速增长,图像样本的语义内容越来越丰富,标签信息也随之越来越复杂.因此图像多标签学习的研究是近年来学术圈和产业界的研究热点之一,涌现了大量表现优异的方法和技术.基于此,本文将对近年来图像多标签学习上的研究成果进行总结.首先,对多标签学习进行简单介绍,并详述其主流方法的分类;随后,针对目前大数据时代的数据特性,总结了多标签学习面临的新的技术难点及其对应的解决方案;最后,在应用层面上介绍了多标签学习在医学、计算机科学等领域的应用实例.

关键词

多标签学习;图像标注;深度学习;大数据

中图分类号 TP391.41;TP183

文献标志码 A

收稿日期 2019-10-15

资助项目 国家自然科学基金(61572503,61872424,6193000388);南京邮电大学高层次人才启动基金(NY218001);模式识别国家重点实验室开放课题(201900015)

作者简介

袁梦奇,男,研究方向为多媒体计算.1337359784@qq.com

鲍秉坤(通信作者),女,博士,教授,博士生导师,主要研究方向为多媒体计算和计算机视觉.bingkunbao@njupt.edu.cn

¹ 南京邮电大学 通信与信息工程学院,南京,210003

0 引言

多标签学习最早起源于机器学习中的单标签学习.图像单标签问题是指通过给定的一组样本特征,来预测该样本所属的类别.比如给定一张动物图片,来预测该动物是否是一条狗.然而,随着大数据时代的到来,数据存储与处理能力得到了飞速的发展,人们对数据分析的要求也开始不断提高.单标签模型很难再刻画大数据时代复杂的现实问题,因此多标签问题诞生了.相比于单标签学习,多标签学习是通过一组样本特征,来预测一张图像对应多个输出标签是否准确.

近年来,涌现了大量表现优异的图像多标签学习方法和技术.本文将对这些研究成果进行归纳和总结,为图像多标签学习的进一步研究提供参考.首先,本文基于对输入样本处理方法的不同,将图像多标签学习分为单例多标签学习(Single-Instance Multi-Label Learning, SIML)和多例多标签学习(Multi-Instance Multi-Label Learning, MIML).单例多标签学习指将一整幅图像看成一个实例,并直接依据图像的整体可视特征(即单一实例的特征),学习其对应的标签.而多例多标签学习是通过将一幅图像根据语义分割为多个区域块(实例),对每个实例进行标签学习.随后,本文针对互联网时代图像数据含有大噪声、数据体量大、关联复杂这三大挑战,总结图像多标签学习面临的技术难点,并介绍对应的解决方案.对于含大噪声的图像,如果直接训练,训练结果将会受到噪声影响,需要首先进行降噪、复原等处理;对于海量数据的问题,将介绍哈希预处理或标签嵌入等方法在降低算法的复杂度上的优势;对于图像间关联复杂方面的挑战,用双标签集或条件贝叶斯框架等方法刻画各图像间的关联,以提高图像的标注性能.最后,介绍多标签学习在多个领域上的应用.在医学中,通过对患者CT图像的多标签学习,完成复杂的疾病监测工作;在无人驾驶中,通过对路况图像信息的多标签学习,有效地帮助汽车处理复杂地形下的自动行驶工作.

本文的结构如下:第1章阐述多标签学习的基本概念及其分类;第2章,将针对多标签学习算法在互联网时代遇到的标签噪声大、数据体量大和标签关联复杂等挑战,整理和总结近年来的研究成果和进展;第3章,讲述多标签学习在医学、计算机科学等领域的实际应用,并对多标签学习未来的发展提出展望.

1 多标签学习概述

本章首先介绍图像多标签学习的基本概念;随后介绍两类经典的多标签学习方法:单实例多标签和多实例多标签,并总结近年来各种分类下多标签学习的研究成果.

1.1 基本概念

图像多标签学习又名图像标注,是依据已知图像内容预测未知标签对象的过程.如图1,对于一张草原照片,可以把标签设为树、大象、狮子、草地等物体^[1],通过输入的图片特征寻找预设的标签是否存在,并进行标注.此外,图像多标签问题还具有三大特点.首先,大部分标签都只与其对应的语义区域有关,这称为标签局部性,如图1中的狮子,只在图像部分区域出现;随后,每个标签间都可能存在某种联系,这称为标签间的相似性,比如在海边图像上,沙滩和大海往往会同时存在;最后,每个标签在不同图像上的对应区域也可以不同,这称为标签内的多样性,比如天空在天气晴朗、多云、下雨时的表现完全不同.

1.2 多标签学习问题的分类

基于对输入样本处理方法的不同,本文将图像的多标签学习方法分为两类:单例多标签和多例多标签.如图2,单例多标签学习指将一整幅图像看成

一个实例,并直接依据图像中的可视特征对标签进行标注的过程^[1],如图1中对大象、狮子、草地的标注.但现实中,图片到标签经常存在一种输入的模糊性,无法根据图像中的可视特征直接标注标签,因此多例多标签学习出现了.多例多标签学习指在一幅图像中提取多个部分,每个部分称作一个实例,并通过这些实例间的语义联系对图像进行多标签学习的过程,如图1中,如果需要标注的标签是非洲,可以先将图像分成树、大象、狮子等实例,再通过学习这些实例完成图像的标注.

图像多标签处理方法大致的演变如图3所示.对于多标签问题的处理,最简单直观的方法是2007年 Tsoumakas 等^[2]提出的将多标签分解为多个单标签问题.但是,这种独立对待每个标签的方法忽略了标签间的相关性,导致训练效果欠佳.于是在2011年 Read 等^[3]提出一种基于分类器链的方法,其通过概率的条件积来捕获标签的依赖性.然而,此方法在处理大量标签时不仅会产生高计算成本,而且其捕获标签间相关性的能力也很有限.随着机器学习等相关技术的快速发展,2013年 Nam 等^[4]开始使用神经网络解决多标签学习,它通过网络内部复杂的神经结构、权重损失函数来优化训练.随着近几年神经网络技术的更进,深度神经网络(Deep Neural Network, DNN)、循环神经网络(Recurrent Neural Net-

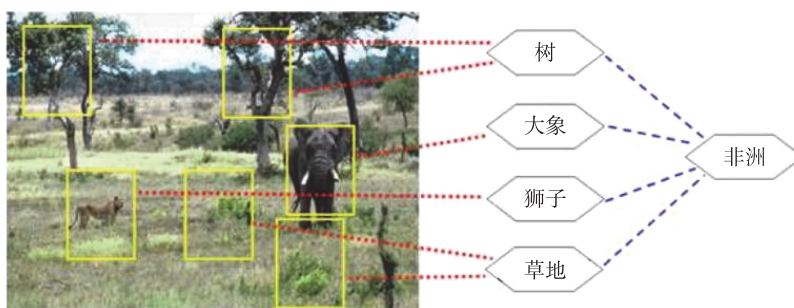


图1 草原图像的多标签学习^[1]

Fig. 1 Multi-label learning of grassland image^[1]

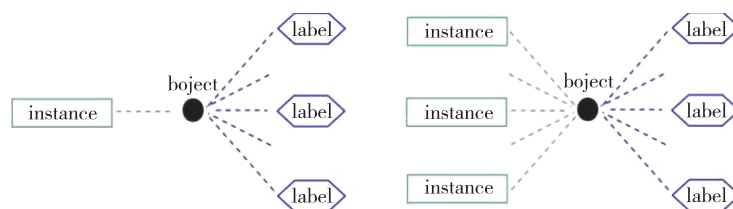


图2 单例多标签和多例多标签^[1]

Fig. 2 Single-instance multi-label and multi-instance multi-label^[1]

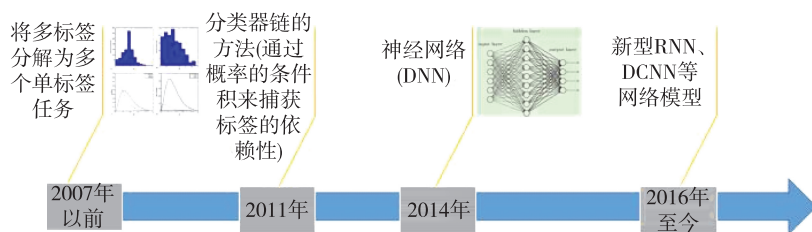


图3 多标签处理方法的演变

Fig. 3 Evolution of multi-label processing

work, RNN)等更为复杂的神经网络开始被用于多标签训练,如基于视觉注意的RNN深度学习框架^[5]、以人类行为为基础的多层次深度卷积神经网络^[6]、区域潜在语义依赖网络模型^[7]等方法。

1.2.1 单例多标签处理方法

单例多标签不需要对图像的输入特征分块处理,所以其模型更简单.而且在图像标注中,很多直观的多标签问题都可由单例多标签处理,比如图1中直接标注图像中的狮子.

对于单例多标签学习,Li等^[8]最先提出一种学习深度二进制编码(DeepBe)的框架,通过将多标签转换为单个标签来处理多标签问题,并且通过集成策略,成功地提高了学习的稳健性.但这种转化过程,对小目标的多标签图像会很难处理.因此,Zhang等^[7]利用了完全卷积定位架构,来定位可能包含多个高度依赖标签的区域.此方法通过标签相关性的使用,让小区域的预测变得容易.但是这类方法并不能很好地处理标签间的依赖关系,为此Markatopoulou等^[6]提出了将多任务学习和多标签学习相结合的处理思路.他们建立一种双层次的深度卷积神经网络框架.在第一个层次上,以多任务学习为基础,提出了一种基于稀疏线性组合来学习每个标签的独有特征;在第二个层次上,以结构化输出学习的思想为基础,考虑标签之间的相关性,其中每个层次都通过标准的卷积层实现,并通过每个词语之间的内在联系提高训练效果.对于建模标签的相关性,Chen等^[5]还提出了一种基于递归神经网络的图像多标签分类模型,通过一个置信度排序的LSTM(Long Short Term Memory)模型对标签的依赖性建模.

除了考虑标签间相关性,He等^[9]提出一种模仿人类行为的强化图像多标签学习方法.该方法允许增强学习代理通过充分利用图像特征和先前的预测结果来按次序预测标签,并在图像从简单到复杂的过程中进行标签化.Li等^[10]则提出了另一种基于成

对排序的多标签图像分类法.其一方面使用了一种新的成对排序损失函数,该函数处处光滑,因而更易于优化;另一方面在模型中加入了一个标签决策模块,用于估计每个视觉对象的最佳置信阈值.此方法在基础的成队排序法上加以优化,也可以达到不错的图像训练效果.此外,在图像多标签学习中降低图的成本也是一个很重要的工作,目前主要的方法是基于最大裕度法来学习边缘权值.但不幸的是这些方法通常对损耗函数(如hamming损耗)提供的约束特别宽松.因此Behpour等^[11]提出了一种基于最小代价图割集的对抗性鲁棒割集的方法,它将学习任务看作是预测者和标签逼近者之间取极大极小值的博弈.

1.2.2 多例多标签处理方法

在实际图像标注问题中,只依据图像的整体可视特征,很难得到准确的语义标签.如图1中的非洲标签,直接看图难以标注,需要先提取图片中大象、狮子等实例,再通过分析这些实例间的语义关系才能分辨此图像是否标注为非洲.这种先将图像看作多个实例,随后通过处理实例间的语义联系对图像进行标注的方法称为多例多标签学习.

针对多例多标签的处理方法,最先Yang等^[12]提出通过结合局部信息的方法来提高图像辨别的能力.此方法首先从每个图像中提取对象建议然后将每一幅图像作为一个包,并将包中提取出的对象建议作为实例,从而把多标签识别问题转化为多类多实例学习问题.但这类方法难以处理多实例与多标签间的联系,因此Ding等^[13]提出了一种同时考虑实例上下文和标签上下文的图像标注上下文感知MIML算法.作者首先引入多个图来描述包中实例之间的复杂关系,并将这些图进行融合,然后利用潜在的概念来包含类标签之间的相互影响.此方法通过利用标签间和实例间的相关性,来达到提升图像分类性能的目的.

然而上述方法无法很好地利用图像的额外信

息,在图像多标签标注的训练中,经常可以获得诸如边界框、图像标题和描述等附加信息,这些信息被称为特权信息.于是 Yang 等^[14]着眼于 MIML 学习中这些特权包的使用,提出一种双流全卷积网络,利用了现有的特权包,而不是难以获得的特权实例,使该系统在实际应用中更具通用性和实用性.而对于图片额外信息有缺失时,Zhu 等^[15]则提出一种利用有效的增广拉格朗日优化算法,用来发现多个新标签.该算法具有一个与包相关的损失项和一个与包无关的聚类正则化项,这使得已知标签和多个新标签可以同时建模,从而达到了在有隐藏的测试数据中预测标签的效果.

此外,Wang 等^[16]提出了另一种多例多标签图像的标注方法:通过开发一个重复记忆注意模块来实现可解释上下文的图像多标签学习.该模块由两个交替执行的部分组成:第 1 部分是一个空间转换层,用于从卷积特征图中自由地定位需要注意区域;第 2 部分是一个 LSTM 子网络,该子网络用于连续地预测所定位区域的语义标记,并同时捕获这些区域的全局相关性.此方法通过学习上下文和可解释的区域的标签相关性,让多标签分类的可辨别性得以提高.

2 多标签学习面临的新挑战

在大数据时代,我们在享受大数据处理带来的方便快捷的同时,也必须得面对大数据图像处理的挑战.大数据下的图像多标签标注有三大特点,分别是标签噪声大、数据体量大和标签关联复杂.首先,如果直接训练标签含噪的图像,训练结果将会受到噪声影响,因此需要进行降噪处理;随后,针对海量数据的问题,需要设计计算复杂度较低的标注算法;最后,应合理利用标签间的关联关系,以提高图像的标注性能.本章针对标签噪声大、数据体量大和标签关联复杂这三大挑战,分别介绍近年来多标签学习的相关研究成果.

2.1 标签噪声大

对于大数据时代的多标签问题,噪声是很常见的,然而直接忽视噪声往往会导致训练效果很差,甚至无法识别图像.因此有效地处理噪声是多标签问题中最先要面对的一个重要挑战.标签大噪声问题,可以分为三类:标签有缺失、标签含噪声差错、标签分布不均衡.下文将依次介绍它们的处理方法.

1) 针对标签有缺失问题.由于在多标签问题中

对多个标签精确地标注是昂贵和困难的,并且拥有完整标签的数据很难获得,所以如何解决含有缺失标签的多标签学习问题变得十分必要.Xie 等^[17]提出一种部分多标签学习(Partial Multi-Label Learning, PML)的方法,作者给每个候选标签一个置信值,以估计它是实例的基本真值标签的可能性.而 Yang 等^[18]则提出一种全局语义描述符和局部语义描述符相连接的语义表示方法,通过一个有效的语义描述符将图像投影到语义空间.全局语义描述部分根据大规模数据集的大量概念来描述“一般的图像是什么”,而局部语义描述符通过对平均特定目标域中每个图像的视觉邻域标签的提取,来描述“图像具体看起来像什么”.最后通过将全局和局部语义描述符相结合,实现了更精确的语义表示.

2) 针对标签含噪声差错问题.在许多实际任务中,人为地标注所有标签成本太高,这时标注器可能会粗略地为每个对象分配一组候选标签,而对于这些候选标签,出现噪声差错也是正常的.对于这种含噪声差错的多标签问题,主要处理方法是部分多标签学习法,而如何具体使用 PML,每位学者也有自己的见解.Sun 等^[19]提出了一种基于低秩稀疏分解的局部多标签学习方法,该方法能够从观测到的标签矩阵中提取地面真值标签矩阵,并在同时通过低秩稀疏分解方案学习预测模型.与此前的 PML 算法相比,该方法能够去除不相关的标签,避免了标签噪声带来的负面影响.而 Fang 等^[20]则提出了另一种基于可信标签启发的部分多标签学习方法:粒子学习,其基本思想是从候选标签集中提取可信的标签来减轻假阳性标签的负面影响,并将其作为可靠的标签信息进行后续模型归纳.

3) 针对标签分布不均衡问题.由于在二分类问题中,通常假设标签正负类别相对均衡,然而实际应用中标签类别不平衡的问题是非常常见的.比如员工体检时不患病人数远超患病人数,产品检测时合格产品数远超不合格产品数等.对于这种分类的样本,由于正例样本出现的概率极低,直接训练会产生严重的过拟合.因此,有必要对这类样本采取特殊的处理方法.Chen 等^[21]提出了一种端到端可训练网络来解决多标记脑肿瘤分割中的标签类间不平衡的问题,通过该网络所具有的肿瘤区域池成分,来减轻非肿瘤区域的影响.然而,在多标签学习问题中,标签缺失和标签不平衡问题经常会同时存在.为了应对这种挑战,Wu 等^[22]提出了一个包含 5 个组成部分

(标签一致性、瞬时水平和类水平标签光滑性,以及2种上下界)的传导学习模型.前3个组件用于将标签信息传播到缺少信息的标签,后2个组件用于处理2种类型的类不平衡问题.

2.2 数据体量大

随着大数据时代的来临,有关多标签数据的存储量得到了飞速的增长.当我们在享受大体量数据所带来的大量重要信息的同时,也不得不面对大体量数据训练困难的问题.如果不做特殊的预处理而直接训练,可能会导致训练复杂度过高,大大影响算法处理的能力.

目前,对于这种体量巨大的数据最有效的预处理方法是基于学习的哈希算法.哈希算法主要用于大体量下的图像检索工作,早期的有监督哈希算法通常将每个图像编码为手工制作的视觉特征向量,然后通过学习哈希函数将这样的特征向量映射为一段二进制代码.近年来,随着深度卷积网络的迅猛发展,一种新兴的有监督哈希方法出现了,它是一种基于深度网络的哈希方法,利用深度卷积网络来学习图像的哈希函数和鉴别特征表示.比如 Huang 等^[23]提出的一种将输入图像编码成二进制码的深层结构,取名叫多标签图像目标位置感知哈希算法,其核心思想是学习一个二值,它可以识别图像对象的近似位置.借助于这些物体的近似位置,来聚焦图像物体的区域,并忽略背景区域,然后将此聚焦的特征表示编码为二进制散列码. Bao 等^[24]则提出一种积累图像语义信息的大规模图像标注方法.作者首先通过多图像分割技术将每个图像编码为区域包;随后,采用局部敏感哈希方法将图像区域分割成多个桶,以此保证相似区域的高碰撞概率;最后,引入稀疏和区域感知的图像相似矩阵有效地提高了图像标注的性能.

除哈希算法外,还有其他一些处理大体量数据的图像多标签学习算法. Gupta 等^[25]从分布语义学的思想中得到启发,提出了一种基于标签嵌入的大规模多标签学习方法,该模型通过有效的基于梯度的方法预测这些嵌入,以便对新输入进行注释. Hu 等^[26]则提出了另一种具有非线性特征转换的网络结构.作者利用海量的数据集和带验证数据的小数据集来学习多标签分类器,也取得了不错的训练结果.

2.3 关联复杂

标签间的相关性是多标签一个重要的特点,通

过标签相关性的合理使用,可以很好地提高图像多标签学习的性能.在过去,处理多标签相关性的方法大致有如下3种:第1种是利用附加信息来运用标签间的相关性^[27];第2种是不考虑数据只考虑标签域中的信息来明确标签间的关联^[28-29];第3种是利用从特征和标签域获得的信息处理^[30-31].不过,近几年又有许多新的利用相关性的思想被提出.其中主要的方法有:双标签集法^[32]、基于协作的多标签学习法^[33]、基于图像特征的处理方法^[34-35]等.

关于双标签集的理论,是一种由 Liu 等^[32]提出,旨在利用集内和集间的标签关系去解决多标签学习问题的方法.作者首先对每个对象给予两组标签来表示相关性,接着提出一种基于 boosting 的框架来解决双标签集问题.其中双标签集的每个组保持一个样本分布,通过使用两个相互重用的基本分类器,让每个标签集利用嵌入在另一个标签集中的信息,并通过样本分布的联合调整,使得一个模型上的错误可由另一个模型弥补,以此来利用集内和集间的标签关系.

协作的多标签学习算法,是由 Feng 等^[33]提出的一种新的多标签学习方法.以前的方法大多将标签相关性作为先验知识,但这可能无法正确刻画标签之间的真实关系.而这种基于协作的多标签学习则可以利用学习到的标签相关性(标签相关性矩阵)来拟合最终预测结果.

关于图像特征的应用,是学者在利用多标签相关性时容易忽略的一点. Li 等^[34]提出了一种条件图形套索的方法来应对这些挑战.这是一种以图像特征的结构和参数学习为基础的贝叶斯框架.将多标签预测问题表述为推理问题,并用平均场变分法求解.关于图像特征的处理, Jing 等^[35]提出另一种方法:标签一致性正则化的多标签字典学习和部分相同标签嵌入法,作者在输入、输出两个空间同时进行多标签学习,从而提高了输入特征空间的特征表示能力和输出标签空间的标签传播能力.

在处理多标签相关性问题时,经常需要面对的另一挑战是缺乏训练数据,而在许多实际应用中,具有可用标签信息的对象数量经常是非常有限的,此时一般的有监督学习算法的性能可能会显著下降.在此基础上, Jing 等^[36]提出了一种半监督奇异值分解法(Singular Value Decomposition, SVD),该方法适用于训练数据量很小的情况下的多标签学习.其主要思想是寻找从特征空间到低秩标签空间的映

射,并通过 SVD 低秩映射出的左/右奇异向量来显式地描述标签/特征分量向量.实验结果表明,此方法也可以在数据量很少的情况下,有效地挖掘和利用特征空间和标签空间中的相关性.

3 多标签学习的应用实例

在实际中,一幅图像往往可同时属于多个分类,需要用多个标签进行标注.因此图像多标签可以很好地为很多实际问题建模.在医学、人工智能、多媒体计算等诸多领域,将目标问题转化为多标签问题常常使得复杂的问题迎刃而解.

在医学领域,很多疾病在多标签的协助下都可以得到一定的处理.比如医学的一大难题:脑肿瘤问题.脑肿瘤虽然不是常见疾病,但它严重危害了患者的健康,导致极高的死亡率.因为手动分割需要高度的专业知识并且非常耗时,所以基于多标签理论的自动脑肿瘤分割技术将大大方便医疗诊断和治疗计划.在将脑肿瘤数据用多标签表示后,通过 Chen 等^[21]提出的一种端到端可训练网络,便可以很好地解决自动脑肿瘤分割问题.Mamani 等^[37]为了更好地处理胸部 CT 扫描图中各器官的自动定位工作,提出一种基于卷积神经网络的多标签结构,通过多池化层的卷积神经网络处理,很好地解决了 CT 图中器官的识别问题.

在人工智能、计算机视觉等方面,多标签技术也有着相当重要的作用.比如在自动驾驶领域,智能汽车的路况识别工作主要是用多标签模型来处理的.为了解决自动驾驶的道路场景识别工作,Chen 等^[38]为驾驶场景数据集提出了一种新的多标签神经网络.该体系结构利用混合标签(同时包括多标签和单标签),其中多标签主要用于多类别预测学习,而单标签则用于对训练过程中需要更仔细处理的小类别进行监督学习,两者协作构建出道路识别模型.Sirirattapol 等^[39]也在路况识别工作中有着独到的见解,他们将多标签卷积神经网络与现有的电视图像结合起来,不需要专门为此类任务设计新的传感器,从而提取出有雨、无雨、晴朗、黑暗、拥挤交通、非拥挤交通、潮湿道路和干燥道路等道路环境情况的组合,不但同时检测多个事件而且降低了成本.

除此之外,多标签还在图像滤波^[40]、视频处理等方面起着重要的作用.虽然在过去十几年科研工作者在图像多标签学习领域取得了许多骄人的成绩,但是在新时期的发展背景下仍然有许多亟待解

决的问题.首先,现实世界是复杂的,图片难以将所有低概率事件完全列出,比如识别被故意涂鸦的汽车.如何以较高的准确性预测受到特殊干扰的图像,始终是一个难题.其次,对于一幅图像上有多个物体的情况,虽然已有许多方法通过语义来刻画物体间的联系,但精度依然有待提高,所以对于物体间关联的刻画依然是图像多标签学习的重点.

4 总结

近年来,随着数据的快速增长,以及数据存储能力的增强,多标签问题的处理工作取得了不错的发展.在医学、计算机等领域,以前很多难以实现的工作,现在都可以通过建立多标签模型,并用神经网络等技术进行处理.本文首先介绍了多标签概念及其分类,接着分别从单例多标签和多例多标签两个角度讲述图像多标签的具体学习方法.随后,本文详细总结了在大数据时代标签噪声大、数据体量大和标签关联复杂这三大挑战下,多标签模型的搭建与处理工作,并构建了一个有效地处理多标签问题的算法指南,同时也为多标签的进一步研究提供了参考.最后,本文简单阐述了多标签在医学、计算机科学等领域的具体现实应用.

参考文献

References

- [1] Zhou Z H, Zhang M L, Huang S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1):2291-2320
- [2] Tsoumakas G, Katakis I. Multi-label classification[J]. International Journal of Data Warehousing and Mining, 2007, 3(3):1-13
- [3] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification[J]. Machine Learning, 2011, 85(3):333-359
- [4] Nam J, Kim J, Gurevych I, et al. Large-scale multi-label text classification: revisiting neural networks[C]// Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2013:437-452
- [5] Chen S F, Chen Y C, Yeh C K, et al. Order-free RNN with visual attention for multi-label classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32:6714-6721
- [6] Markatopoulou F, Mezaris V, Patras I. Implicit and explicit concept relations in deep neural networks for multi-label video/image annotation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(6):1631-1644
- [7] Zhang J J, Wu Q, Shen C H, et al. Multilabel image classification with regional latent semantic dependencies[J].

- IEEE Transactions on Multimedia, 2018, 20 (10) : 2801-2813
- [8] Li C H, Kang Q, Ge G J, et al. DeepBE: learning deep binary encoding for multi-label classification [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016: 744-751
- [9] He S Y, Xu C, Guo T Y, et al. Reinforced multi-label image classification by exploring curriculum [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32: 3183-3190
- [10] Li Y C, Song Y L, Luo J B. Improving pairwise ranking for multi-label image classification [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1837-1845
- [11] Behpour S, Xing W, Ziebart B D. ARC: adversarial robust cuts for semi-supervised and multi-label classification [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018: 1905-1907
- [12] Yang H, Zhou J T, Zhang Y, et al. Exploit bounding box annotations for multi-label object recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 280-288
- [13] Ding X M, Li B, Xiong W H, et al. Multi-instance multi-label learning combining hierarchical context and its application to image annotation [J]. IEEE Transactions on Multimedia, 2016, 18(8) : 1616-1627
- [14] Yang H, Zhou J T Y, Cai J F, et al. MIML-FCN+: multi-instance multi-label learning via fully convolutional networks with privileged information [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5996-6004
- [15] Zhu Y, Ting K M, Zhou Z H. Discover multiple novel labels in multi-instance multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31: 2977-2984
- [16] Wang Z X, Chen T S, Li G B, et al. Multi-label image recognition by recurrently discovering attentional regions [J]. IEEE International Conference on Computer Vision (ICCV), 2017: 464-472
- [17] Xie M K, Huang S J. Partial multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32: 4302-4309
- [18] Yang H, Zhou J T, Cai J F. Improving multi-label learning with missing labels by structured semantic correlations [M] // Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 835-851. DOI: 10.1007/978-3-319-46448-0_50
- [19] Sun L J, Feng S H, Wang T, et al. Partial multi-label learning by low-rank and sparse decomposition [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 5016-5023
- [20] Fang J P, Zhang M L. Partial multi-label learning via credible label elicitation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 3518-3525
- [21] Chen X, Liew J H, Xiong W, et al. Focus, segment and erase: an efficient network for multi-label brain tumor segmentation [M] // Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 674-689. DOI: 10.1007/978-3-030-01261-8_40
- [22] Wu B Y, Lyu S W, Ghanem B. Constrained submodular minimization for missing labels and class imbalance in multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30: 2229-2236
- [23] Huang C Q, Yang S M, Pan Y, et al. Object-location-aware hashing for multi-label image retrieval via automatic mask learning [J]. IEEE Transactions on Image Processing, 2018, 27(9) : 4490-4502
- [24] Bao B K, Ni B B, Mu Y D, et al. Efficient region-aware large graph construction towards scalable multi-label propagation [J]. Pattern Recognition, 2011, 44 (3) : 598-606
- [25] Gupta V, Wadbude R, Natarajan N, et al. Distributional semantics meets multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 3747-3754
- [26] Hu M Y, Han H, Shan S G, et al. Multi-label learning from noisy labels with non-linear feature transformation [C] // Asian Conference on Computer Vision, 2019: 404-419
- [27] Bengio S, Weston J, Grangier D. Label embedding trees for large multi-class tasks [C] // NIPS'10 Proceedings of the 23rd International Conference on Neural Information Processing Systems, 2010: 163-171
- [28] Dembczynski K, Cheng W, Hullermeier E. Bayes optimal multi-label classification via probabilistic classifier chains [C] // International Conference on Machine Learning (ICML), 2010: 279-286
- [29] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23 (7) : 1079-1089
- [30] Chen Y N, Lin H T. Feature-aware label space dimension reduction for multi-label classification [J]. Advances in Neural Information Processing Systems, 2012, 2: 1529-1537
- [31] Li X, Zhao F P, Guo Y H. Conditional restricted boltzmann machines for multi-label learning with incomplete labels [C] // Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, 2015: 635-643
- [32] Liu C, Zhao P, Huang S J, et al. Dual set multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32: 3635-3642
- [33] Feng L, An B, He S. Collaboration based multi-label learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 3550-3557
- [34] Li Q, Qiao M Y, Bian W, et al. Conditional graphical lasso for multi-label image classification [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2977-2986
- [35] Jing X Y, Wu F, Li Z Q, et al. Multi-label dictionary learning for image annotation [J]. IEEE Transactions on Image Processing, 2016, 25(6) : 2712-2725
- [36] Jing L P, Shen C Y, Yang L, et al. Multi-label

- classification by semi-supervised singular value decomposition [J]. IEEE Transactions on Image Processing, 2017, 26(10):4612-4625
- [37] Mamani G E H, Setio A A A, Ginneken B V, et al. Organ detection in thorax abdomen CT using multi-label convolutional neural networks [C] // SPIE Medical Imaging, 2017:1013416
- [38] Chen L, Zhan W J, Tian W, et al. Deep integration: a multi-label architecture for road scene recognition [J]. IEEE Transactions on Image Processing, 2019, 28(10): 4883-4898
- [39] Sirirattapol C, Nagai M, Witayangkurn A, et al. Bangkok CCTV image through a road environment extraction system using multi-label convolutional neural network classification [J]. ISPRS International Journal of Geo-Information, 2019, 8(3): 128
- [40] Dai L Q, Yuan M K, Li Z C, et al. Hardware-efficient guided image filtering for multi-label problem [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:4905-4913

A survey of image multi-label learning

YUAN Mengqi¹ BAO Bingkun¹

¹ School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003

Abstract With the fast growing number of images, especially the user-generated ones, the semantic content of images become richer, and labels become more complex. Therefore, the study on image multi-label learning is one of the hot research areas in both academia and industry, and a large number of efficient methods have emerged in recent years. This paper surveys the existing work on image multi-label learning in recent years. Firstly, we briefly describe the concept of multi-label learning and introduce two types of methods, that is, single-instance multi-label learning and multi-instance multi-label learning. Then, we summarize three challenges on multi-label learning caused by the big data characteristics, and provide related work which can handle these challenges. Finally, we elaborate two applications on image recognition and automatic drive to show that multi-label learning techniques can be effective for many application scenarios.

Key words multi-label learning; image annotation; deep learning; big data