

图像多标签分类综述

张逸菁 宋雅捷 王云飞
新疆大学

DOI:10.12238/acair.v2i2.7344

[摘要] 多标签图像分类是计算机视觉领域的重要研究方向,它的目标是对同时包含多个对象的图像进行分类,并准确识别出各个类别。以往在对图像进行分类时,通常选取单个对象作为其标签,但考虑到现实生活中一张图像往往存在多个对象,多标签图像分类也越来越受到关注。因此对传统机器学习以及深度学习中的多标签图像分类方法进行研究分析。同时对多标签图像未来要进行的工作进行展望,希望对多标签图像分类问题的发展与进步起到一定推进作用。

[关键词] 多标签图像分类; 机器学习; 深度学习

中图分类号: TP181 文献标识码: A

Overview of Image Multi label Classification

Yijing Zhang Yajie Song Yunfei Wang
Xinjiang University

[Abstract] Multi label image classification is an important research direction in the field of computer vision. Its goal is to classify images containing multiple objects simultaneously and accurately identify each category. In the past, when classifying images, a single object was usually selected as its label. However, considering that in real life, an image often has multiple objects, multi label image classification has also received increasing attention. Therefore, research and analyze multi label image classification methods in traditional machine learning and deep learning. At the same time, we look forward to future work on multi label images, hoping to play a certain role in promoting the development and progress of multi label image classification problems.

[Key words] Multi label image classification; Machine learning; Deep learning

引言

多标签图像广泛地存在于我们的生活中,用于表示复合对象元素之间的复杂关系。近年来,对多标签图像数据的研究是学术界的一个热点,即给定一组图像,多标签图像分类的目标是学习图像和对应所有类别标签的映射关系,并预测未知图像的类别标签。多标签图像分类是一个重要的数据挖掘任务,可以应用在很多领域,例如生物信息学中,通过蛋白质图像分类判断蛋白质是不是酶,是不是具有对某种疾病的治疗能力^[1,2]。当前还没有同时包括基于传统机器学习和基于深度学习的多标签图像分类研究综述。为了方便更多的研究人员,本文梳理总结了多标签图像分类的各类研究方法。

1 图像多标签分类问题的定义及目前面临的挑战

1.1 图像多标签分类问题的定义。图像多标签分类问题是计算机视觉领域的一个重要研究方向。与单标签分类不同,图像多标签分类的目标是识别出一张图像中所包含的多个标签。在我们对实际上的图片进行分类时,一张图片也往往包含着多个对象,因此,图像多标签分类也与我们实际面对的问题更为契合。

1.2 目前面临的挑战。(1)数据量和标注困难。图像多标签分类任务通常需要大量的带有多个标签的图像数据进行训练。然而,获取带有多个标签的大规模数据集并进行正确标注是一项昂贵且耗时的任务。(2)标签间的相关性。由于图像多标签分类问题中一张图像存在多个标签,且这些标签往往存在着相关关联性,因此,对于图像多标签分类任务,为了提高模型识别的准确性,让模型能够捕捉到标签之间的相关性非常重要。(3)标签不平衡问题。图像多标签分类任务中,不同类别之间可能存在着不平衡的问题。某些类别可能出现频率较高,而其他类别可能出现频率较低,这可能会导致模型倾向于预测频率较高的类别,而忽略其他类别。(4)处理大规模图像数据的效率。多标签图像分类的可能性随着图片中标签类别的增加呈指数级增长,在现有的硬件基础上会加剧训练的负担和时间成本,如何有效的降低信息维度是面临的最大挑战。

2 基于传统机器学习的图像多标签分类算法

传统的机器学习进行多标签图像分类主要需要进行特征提取及特征分类两个步骤,在图像分类中对于分类模型的使用主要有以下几种。

2.1 问题迁移。通过某种方式将多标签数据集转换为单标签数据集, 进而通过单标签图像分类的方式进行分类。

(1) 二元关联(binary relevance)。二元关联在图像多标签分类领域可以用于解决标签之间的相互关系和相关性。这种方法可以帮助改善多标签分类的准确性, 因为标签之间往往存在相互关联。二元关联可以用于学习相关性、预测相关标签、处理层次结构和权重学习等任务, 以更精确地为图像分配多个标签。(2) 分类器链(classifier chains)。分类器链方法将原始的多标签分类问题分解为多个独立的二分类子问题, 每个子问题对应于一个标签, 因此可以构建多个二分类器来预测每个标签的存在或缺失。每个二分类器的输出都作为下一个分类器的输入, 每个分类器都根据前一个分类器的输出和图像特征进行训练和预测考虑标签之间的相关性, 以序列化的方式处理这些标签, 从而提高多标签分类的性能。

2.2 传统算法改进。

2.2.1 KNN。k近邻算法的思想非常易于理解, 于一个样本可以取到距离它最近的M个样本, 然后得到这M个样本的标签集合, 进而通过最大算法认为, 如果一个样本在其特征空间中的k个距离最临近的样本大多数属于某一个类别, 那么这个样本也应该属于这个类别。在单标签识别中与一个样本在特征空间中距离越近的样本, 他们的标签是相同的概率就越大^[3]。

根据这个思想, 将k近邻算法应用于多标签分类领域, 对后验概率准则来确定这些^[4]。KNN算法虽然简单易懂, 但也存在着没有考虑标签相关性等图像多标签分类普遍存在的问题, 因此钱龙团队提出了S-KNN, 该算法考虑了标签之间的相关性, 提升了算法识别的准确性^[5]。

2.2.2 SVM(向量机)。支持向量机(support vector machines, SVM)是一种二分类模型, 它将实例的特征向量映射为空间中的一些点, SVM的目的就是想要画出一条线, 以“最好地”区分这两类点, 以至如果以后有了新的点, 这条线也能做出很好的分类。SVM适合中小型数据样本、非线性、高维的分类问题^[6]。

2.2.3 Multi-label decision tree。多标签决策树(Multi-label Decision Tree)是决策树算法的一种扩展, 专门用于处理多标签分类问题。在多标签分类中, 每个实例可以同时与多个标签相关联, 这与传统的二元分类或多类分类不同, 后者每个实例仅与一个标签相关联。多标签决策树用于在需要为单个数据点预测多个标签或类别的情况^[7]。

2.3 集成方法。集成方法是一种机器学习技术, 它通过结合多个基本模型的预测结果, 以提高整体性能和泛化能力。这些基本模型可以是不同的算法或是同一算法的不同实例, 集成它们的预测结果可以减小单个模型的偏差和方差, 从而提高模型的鲁棒性和性能^[8]。

3 基于深度学习的图像多标签分类算法

近年来深度学习对于图像多标签分类领域产生了重大影响, 深度学习模型在图像多标签分类任务上展现了其出色的性能, 避免了手动提取特征的繁琐性, 并且在准确度等多个方面都超

越了传统的机器学习。

3.1 CNN。卷积神经网络是一类专门用于处理具有网格结构的数据的深度学习模型, 由卷积层、池化层、全连接层构成。卷积层使用一系列滤波器(卷积核)对输入数据进行卷积, 检测输入数据中的不同特征; 池化层降低特征图的空间维度、减少计算量、提升模型稳定性; 全连接层将经过卷积与池化提取出的特征映射到输出类别, 每个类别都与上一层的节点相接, 形成全连接^[9]。如: EfficientNet模型, 它在图像分类任务上表现优秀, 通过使用复合缩放系数(compound scaling)和神经网络结构搜索(neural architecture search)的方法, 在提高模型性能的同时, 控制了模型参数和计算复杂度的增长, 具有较强的泛化能力^[10]。

3.2 深度学习模型。这类模型专门设计用于图像多标签分类任务。它们使用深度卷积神经网络, 通常具有多个卷积层和全连接层。模型的最后一层采用sigmoid函数作为激活函数, 以输出每个可能标签的概率。这种模型可以同时预测图像中存在的多个标签。

其中由于深度学习需要大量的标注数据和计算资源, 迁移学习成为一种常用的图像分类方法。迁移学习通过使用在大规模图像数据集上预训练的模型, 如: ImageNet上的预训练模型ResNet模型, 通过微调即可使其应用在目标的图像多标签分类任务中; BERT模型, 由Google提出的自然语言处理模型, 将图像用自然语言表述以后同样可以使用该模型进行处理^[11]。

3.3 注意力机制。注意力机制模型通常包括卷积神经网络和注意力机制模块, 注意力机制模块能够学习到图像中重要区域的权重, 从而加强这些区域对标签分类的影响。

目前主流的注意力机制模型, 如: Transformer模型, 它是一种基于自注意力机制的深度学习模型, 在处理序列数据和实现机器学习任务时表现出了更优秀的性能; 基于Transformer提出的ViT模型可以学习到图像中不同位置的关系和上下文信息从而实现图像多标签分类^[12]。

4 图像多标签分类的应用场景与未来的研究方向

图像多标签分类问题在自动驾驶、内容检索等众多应用领域展现了其不可或缺的价值。通过赋予单一图像多个标签, 不仅能更准确地捕捉到图像内容的丰富性, 还能提供更加细致和全面的信息解读。然而, 这一领域仍然有很多需要注意的问题和值得探索的方向。

4.1 应用场景。(1) 自动驾驶。在车辆实时识别与自动驾驶中, 图像处理起着关键作用。通过图像多标签分类, 可以将车辆图像中的不同特征进行分割, 并为每个特征打上相应的标签。这样, 系统能够实现对车辆的实时识别, 为自动驾驶车辆提供精准的环境感知, 确保驾驶决策的准确性和安全性^[13]。这种方法不仅提高了自动驾驶系统对道路状况的理解能力, 也为车辆在复杂交通环境中的安全行驶提供了有力支持。(2) 医学影像分析。医学影像具有复杂的视觉特征, 同时还伴随着多种标签的天然属性。在传统医疗领域中, 医生往往依赖个人经验来做出诊断判断, 但这种方法存在准确性不高且难以形成统一的标准的问题。因此, 图像多标签分类技术可以通过为不同病变对应的医学图

像添加标签^[14],更好地模拟医生对医学影像的认知过程,提高医学影像的分析准确性和一致性,为医疗诊断提供更可靠的支持。(3)生物学。在生物学中,蛋白质之间并非完全独立,多种蛋白质共同构成单个细胞,因此研究中需要同时识别多种蛋白质,传统的单标签分类模型无法胜任^[15],图像多标签分类常见应用于蛋白质细胞图谱构建等问题中。(4)电力系统分析。电力系统的电能质量扰动逐渐复杂化与多样化,传统的单标签分类方法无法解决标签集外的复合扰动^[16],因此将一维时域扰动信号转换为二维可识别图像,采用图像多标签分类来识别检测电力系统中的复合扰动。

4.2未来研究方向。图像多标签分类问题在图像的分类任务上已经有了很多研究成果,但是该领域仍然有很多需要注意的问题和值得继续探索的研究方向。

(1)深度学习模型的优化。如何进一步优化深度学习模型以提高分类精度和效率仍然是一个重要的研究方向。可以从网络结构设计方面设计更深的网络结构,引入更复杂的非线性变换等;从训练方法方面采用更有效的损失函数,使用更先进的正则化技术,或者引入更复杂的数据增强策略等;除此之外优化算法也是优化模型的重要手段,可以研究更有效的梯度下降算法或者更先进的优化策略。(2)大规模数据处理。随着互联网的发展,图像数据量呈指数级增长,如何在大规模数据上进行有效的多标签分类将是另一个重要的研究方向。未来需要研究新的数据预处理方法,设计更有效的数据清洗和标注策略,使用更先进的特征提取和降维技术等。因此并行和分布式计算也是未来研究的重要方向,可以研究更有效的并行计算框架,设计更高效的分布式计算策略,引入更先进的硬件设备。(3)标签关联性研究。在多标签分类问题中,不同标签之间往往存在一定的关联性,如何利用这种关联性来提高分类效果将是未来的一个重要研究方向。未来需要研究标签关联性的建模方法,设计更复杂的标签关联性模型,使用更先进的关联性度量方法,引入更复杂的图模型等。基于标签关联性的分类方法也是未来研究的重要方向,可以研究更有效的关联性驱动的分类算法,设计更高效的关联性优化策略,引入更先进的决策理论。(4)弱标签和未知标签学习。在实际应用中,往往存在大量的未知标签数据和弱标签数据,针对这个问题未来需要研究新的弱标签和未知标签学习方法,设计更有效的弱标签和未知标签模型,使用更先进的弱标签和未知标签优化算法,引入更复杂的半监督学习策略等或者迁移学习理论。

5 总结

图像多标签分类在计算机视觉领域一直都非常重要,有诸多研究成果不断涌现。文章从传统的机器学习以及迅速发展的深度学习两方面介绍图像多标签分类的主要方法。可以看出卷积神经网络作为主流模型,在图像特征提取方面取得了显著的成果,多种深度学习模型的提出为图像多标签分类给出了诸多解决方案。但是,图像多标签分类研究依旧面临着如:标签之间的相关性、样本不平衡、大规模数据集等挑战,因此研究者应该设计更有效的模型算法、更高效的优化策略,引入更先进的决

策理论,考虑深度学习模型与传统机器模型有效耦合等方案。未来图像多标签分类的研究方向将会是在关注模型性能的同时,进一步提升模型对复杂场景的处理能力,不断深入该领域的研究,将会对实际应用场景产生重要的影响。

[基金项目]

新疆大学国家级大学生创新创业训练计划资助,项目编号:202210755094。

[参考文献]

- [1]Borgwardt KM,Kriegel HP,Vishwanathan SVN,Schraudolph NN.Graph kernels for disease outcome prediction from protein-protein interaction networks. Pacific Symposium on Biocomputing,2007:4-15.
- [2]Fout AM. Protein interface prediction using graph convolutional networks[MS. Thesis]. Colorado:ColoradoStateUniversity,2017.6530-6539.
- [3]Tenindra Abeywickrama,Muhammad Aamir Cheema,and David Taniar.2016.K-nearest neighbors on road networks:a journey in experimentation and in-memory implementation.Proc. VLDB Endow.9,6(January 2016),492 - 503.https://doi.org/10.14778/2904121.2904125.
- [4]李峰,苗夺谦,张志飞,等.基于互信息的粒化特征加权多标签学习k近邻算法[J].计算机研究与发展,2017,54(05):1024-1035.
- [5]钱龙,赵静,韩京宇,等.基于标签相关性的K近邻多标签学习[J].计算机工程,2022,48(06):73-78+88.
- [6]孙浩荣.基于SVM的食物图像分类算法的研究[D].西安电子科技大学,2016.
- [7]杨朋波,桑基韬,张彪.面向图像分类的深度模型可解释性研究综述[J].软件学报,2023,34(01):230-254.
- [8]张思青.面向细粒度图像分类的知识蒸馏方法研究[D].北京邮电大学,2022.
- [9]李爱萍.基于CNN的多示例多标签图像分类[D].西北大学,2021.
- [10]Tan,M.,& Le,Q.V.(2019).EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.ArXiv,abs/1905.11946.
- [11]Zhong J, Xuan Z, Wang K, Cheng Z.A BERT-Span model for Chinese named entity recognition in rehabilitation medicine.PeerJ Comput Sci.2023 Aug21;9:e1535
- [12]Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J].2020.
- [13]曹艳秋.面向车载视频的多标签分类方法研究[D].南京师范大学,2021.
- [14]魏琦.基于知识图谱的医疗图像多标签分类及未知疾病检测[D].云南大学,2022.