

复杂数据背景下的多标记学习研究进展

杜国栋¹, 王 敖¹, 路鹏伟¹, 叶倩芝², 张 佳²⁺

1. 燕山大学 信息科学与工程学院, 河北 秦皇岛 066004

2. 暨南大学 信息科学技术学院, 广州 510632

+ 通信作者 E-mail: jiazhang@jnu.edu.cn

摘要: 多标记学习是机器学习领域中的一种重要学习范式, 其在许多实际应用中展现出先进的性能。以大规模、信息不完备、多维度和动态演化等为主要特征的复杂数据给多标记学习带来了重大挑战。因此, 有必要依据复杂数据背景下不同学习任务的特点对多标记学习方法进行更为精细的设计。为此, 详细分析了多标记数据复杂性的来源并系统综述了复杂数据背景下的多标记学习研究进展; 介绍了多标记学习的问题定义、学习形式和发展趋势。围绕着“标记”“特征”和“样本”三个层面对多标记数据的复杂性进行探讨。对复杂数据背景下的多标记学习算法进行了详细的归纳、阐述和分析; 总结了面临的挑战和未来可能的研究方向。

关键词: 多标记学习; 标记复杂性; 特征复杂性; 样本复杂性

文献标志码: A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.2507-0192

Survey on Multi-Label Learning Under Complex Data Environment

DU Guodong¹, WANG Ao¹, LU Pengwei¹, YE Qianzhi², ZHANG Jia²⁺

1. School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

2. College of Information Science and Technology, Jinan University, Guangzhou 510632, China

Abstract: Multi-label learning is an important learning paradigm in the field of machine learning, which shows advanced performance in many applications. Data with the main characteristics of large-scale, incomplete information, multiple dimensions and dynamic evolution are bringing many challenges to multi-label learning. It is necessary to carry out a more refined design of the multi-label learning method according to the characteristics of different learning tasks under complex data environment. Therefore, this paper analyzes the source of the complexity of multi-label data and reviews the research progress of multi-label learning under complex data environment. Firstly, the paper introduces the problem definition, learning form and development trend of multi-label learning in detail. Next, the complexity of multi-label data from three levels: label, feature and sample is discussed. Then, the summarized, elaborated and analyzed multi-label learning methods under complex data environment are provided. Finally, the challenges and the possible future application directions of multi-label learning are summarized.

Key words: multi-label learning; label complexity; feature complexity; sample complexity

多标记数据具有特殊的表现形式, 数据中的每个样本通常由一个示例表示, 同时该样本具有多种语义信息, 也就是与多个类别标记相关^[1]。这种形式的数据在真实世界中广泛存在, 例如在文档分类中, 一篇体坛新闻报道可能隶属于“体育”主题, 还可能隶属于“游泳”或者“跳水”等主题^[2]; 在图像标注中, 一副图像可能同时属

于“日落”“海”和“沙滩”等多个类别^[3]; 在生物信息学中, 单个基因可能同时具有“新陈代谢”“蛋白质合成”和“转录”等多种功能^[4]。传统的监督学习算法通常为每个标记分别构建一个二类分类器来预测未知样本的相关标记。然而, 当面对多标记数据时, 这种方式存在诸多弊端。标记过多会导致区分一个标记的训练样本增加, 若

基金项目: 河北省自然科学基金(F2025203073); 河北省教育厅科学研究项目(QN2025001); 燕山大学基础创新科研培育项目(2024LGQN004); 河北省创新能力提升计划项目(22567626H); 中央高校基本科研业务费专项资金(21625110)。

作者简介: 杜国栋(1992—), 男, 博士, 讲师, CCF专业会员, 研究方向为机器学习、数据挖掘; 王敖(2002—), 男, 硕士研究生, 研究方向为机器学习、数据挖掘; 路鹏伟(2003—), 女, 硕士研究生, 研究方向为数据挖掘、机器学习; 叶倩芝(1996—), 女, 博士研究生, 研究方向为机器学习、数据挖掘; 张佳(1991—), 男, 博士, 讲师, CCF专业会员, 研究方向为机器学习、数据挖掘。

收稿日期: 2025-07-14 **修回日期:** 2025-10-20 **文章编号:** 1002-8331(2026)11-0041-21

某个标记的正类样本很少则极易使模型性能下降^[5]。假设标记之间相互独立,为每个标记独立构建二类分类器,又会引发计算复杂度增加的问题,尤其是在标记数量较多时,这种方法的效率极为低下。因此,学习算法必须具备高效的计算能力,以应对标记数量的增长。为此,多标记学习应运而生。传统监督学习任务中每个样本只有一个相关的标记。而多标记学习任务通常更具挑战性,因为每个样本有多个相关联的标记,且标记之间可能存在相互关系,如图1所示。

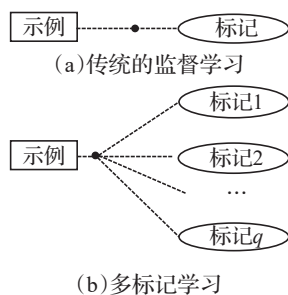


图1 两种监督学习框架

Fig.1 Two supervised learning frameworks

近年来,多标记学习取得了长足的发展,并在图像识别、文本分类、多任务学习、基因分析、推荐系统等多个领域取得了巨大成功。一个重要的原因是其能自然地表达学习对象丰富的语义信息,并且能探索多标记数据的内在性质来建模,从而可以捕获标记之间的关系来完善标记自身的学习能力^[6]。多标记学习通过对标记之间的关系进行有效建模,进而显著提升预测能力。例如,在医疗诊断中,“咳嗽”“发烧”和“肺炎”等标记存在特定关联,多标记学习能利用这些关联提高诊断准确性。此外,多标记学习在处理高维标记时设计了高效的联合优化框架,能保持较低的计算复杂度,解决了传统方法在标记数量激增时的效率瓶颈。随着大数据时代的不断发展,各行各业都面临着大规模数据的快速增长,促使多标记学习的应用需求不断扩大,同时也激发了相关技术的深入研究。例如,在社交网络分析中,一条热门话题可能瞬间衍生出多个相关标记,多标记学习能快速适应这种动态变化,及时准确地进行分类和预测。综上所述,多标记学习不仅丰富了信息表达的手段,也提供了应对复杂、多变数据的有效工具,这充分体现了其研究的必要性和广阔的应用前景。

此外,大规模、信息不完备、多维度 and 动态演化等复杂类型的多标记数据对现有计算方法带来巨大挑战。例如,在实时数据分析场景中,数据的动态演化要求多标记学习算法能够快速适应数据变化,及时更新模型以保证预测的准确性;而在一些实际应用中,数据可能存在缺失和噪声等问题,进一步增加了多标记学习的难度。解决这些问题对于推动多标记学习在各个领域的广泛应用具有重要的现实意义,能够为实际问题的解决提供更

有效的方法和技术支持。为此,本文从多标记学习适用复杂数据的特点对相关算法进行综述。主要贡献如下:

(1) 现有多标记学习综述^[7-11]侧重于多标记学习框架的描述,主要考虑利用多标记学习区别于传统监督学习的特点来对算法进行分类;而本文主要是依据多标记数据计算的复杂性来对适用于不同复杂数据情形的学习算法进行分类,弥补已有综述工作的不足。

(2) 针对复杂数据背景下的多标记学习研究进展,系统地标记、特征和样本三个层面调研了现有多标记学习算法的研究现状,并详细分析了可能存在的问题和未来的研究方向。

(3) 本文以中医健康状态辨识和药物研发为例,对多标记学习未来的应用方向和发展趋势进行展望。

为此,本文首先对多标记学习进行概述,进而从标记、特征和样本三个层面对复杂数据背景下多标记学习面临的挑战进行分析。然后,详细介绍了适用于不同复杂数据处理的算法研究现状,展望了多标记学习未来的研究方向。最后,调研了包括中医状态辨识和药物研发等多标记学习潜在应用方向。

1 多标记学习概述

1.1 多标记学习任务

为了形式化定义多标记学习问题,本文使用以下符号和术语。假设一个训练数据集有 n 个样本,每个样本可能与标记集合 $L = \{l_1, l_2, \dots, l_q\}$ 中的 q 个标记相关联,则该数据集可以表示为 $D = \{(x_i, Y_i) | 1 \leq i \leq n\}$ 。其中, x_i 为第 i 个示例,该示例由一个 d 维的特征向量表示,则这 n 个示例组成数据集 D 的特征空间可以表示为 $X \in \mathbb{R}^{n \times d}$ 。 $Y_i = \{y_{i1}, y_{i2}, \dots, y_{iq}\}$ 为 x_i 的标记向量,包含与 x_i 相关的标记子集, n 个样本对应的标记向量组成数据集 D 的标记空间 $Y \in \mathbb{R}^{n \times q}$ 。对于 Y_i 中的任意元素 $y_{ij} (1 \leq j \leq q)$, $y_{ij} \in \{-1, 1\}$ 。若 x_i 隶属于标记 l_j , $y_{ij} = 1$, 否则 $y_{ij} = -1$ 。假设空间表示从输入空间到输出空间的映射函数的集合。多标记学习的目标是从假设空间中选择一个映射函数 $f: X \rightarrow 2^Y$, 使其准确地预测每个样本的多个标记。定义 $U = \{(x_i, Y_i) | 1 \leq i \leq t\}$ 为包含 t 个样本及其对应标记的集合,希望得到的分类器 f 能够对数据集 U 中的未见示例做出尽可能准确的预测。

1.2 多标记学习评价指标

在多标记学习中,由于每个样本可以有多个标记,因此传统的单标记学习的评价指标不再适用。为了评估多标记学习算法的性能,一些特定的评价指标被引入。这些指标在多标记学习中应用最广泛^[12]。其中, Hamming loss、Ranking loss、Coverage 和 One-error 越小越好; Average precision 越大越好。这五个指标均是基

于样本的评价准则^[13],值得说明的是,还有一些基于标记的评价准则,如Macro-F1、Micro-F1、Macro-AUC、Micro-AUC是值均越大越好。此外,Hamming loss、Macro-F1、Micro-F1、Macro-AUC、Micro-AUC用于评价所有标记的分类错误;而Ranking loss、Coverage、Average precision用于评价标记相对位置的排序错误;One-error则用于评价排在最前端的标记的分类错误。这些指标提供了多个方面的评价,覆盖了多标记学习中标记多样性、样本多标记等特点。表1为常用的多标记学习评价指标。在选择评价指标时,需要考虑任务的特性以及模型对任务的适应程度。另外,各指标之间的关系尚不明确,优化其中一个指标是否意味着优化其他指标不得而知。因此,一般都是选取多个不同指标作为评价标准。

1.3 标记关系利用

在多标记学习中,标记之间的关系可以对模型的性能产生重要影响,这也是多标记学习不同于传统机器学习的显著特点^[20]。标记关系的利用使得模型能够更好地建模标记之间的相关性和依赖关系。在高维度的多标记学习任务中,标记之间的关系可以帮助降低维度灾难的影响^[21]。通过有效地选择和利用与任务相关的标记关系,可以减少特征空间的维度,提高模型的泛化能

力。标记关系的利用有助于提高模型在未见样本上的泛化能力^[22]。最初的方法是将每个标记视为一个独立的二分类问题。每个标记都有一个独立的分类器来判断它是否存在,分类器可以选择传统的监督学习方法,其中Boutell等^[23]使用SVM实现多标记分类的任务,Clare等^[24]则采用决策树对其进行分类。传统的方法虽然简单直接,但是忽略了标记之间的相互关系。因此,本文从标记关系利用的角度可以将多标记学习方法分为五个类别,即问题转化方法、从数据中直接学习的方法、优化损失函数的方法、利用标记层次结构的方法以及利用标记局部关系的方法。

(1)问题转化方法:这类方法是将多标记问题转换为多个单标记问题,以便使用传统的单标记学习算法进行分类。Random k -Labelsets是这种策略的代表性方法^[25],该算法每次从训练数据的标记集合中随机选择 k 个标记的子集,然后每个子集单独训练一个分类器,最后对生成的分类器结果进行聚合。而分类器链Classifier chains算法^[26]则是将标记按照一定的顺序链接起来,每个分类器的输出不仅取决于样本的特征,还取决于链前面标记的输出,因此可以通过链结构捕捉标记之间的关联性。Classifier chains在多标记分类任务上具有不错

表1 多标记学习评价指标
Table 1 Evaluation metrics of multi-label learning

评价指标	计算公式	含义
Hamming loss ^[14]	$hloss(f, \mathcal{U}) = \frac{1}{l} \sum_{i=1}^l f(x_i) \oplus Y_i _1$	用于考察实际标记与预测标记不匹配的平均值,同时考虑了标记之间的不同权重
Ranking loss ^[15]	$rloss(f, \mathcal{U}) = \frac{1}{l} \sum_{i=1}^l \frac{ \{(l_j, l_k) f_j(x_i) \leq f_k(x_i), (l_j, l_k) \in Y_i \times \bar{Y}_i\} }{ Y_i \bar{Y}_i }$	用于考察模型的排序质量,即正确标记的排名在所有标记中的平均位置
Coverage ^[9]	$coverage(f, \mathcal{U}) = \frac{1}{q} \left(\frac{1}{l} \sum_{i=1}^l \max_{l_k \in Y_i} rank(x_i, l_k) - 1 \right), rank(x_i, l_k) = \sum_{j=1}^q \mathbb{I}[f_j(x_i) \geq f_k(x_i)]$	用于考察平均每个样本需要多少个标记才能正确覆盖实际标记
Average precision ^[16]	$aveprec(f, \mathcal{U}) = \frac{1}{l} \sum_{i=1}^l \frac{1}{ Y_i } \sum_{l_j, l_k \in Y_i} \frac{L_i = \{l_j rank(x_i, l_j) \leq rank(x_i, l_k)\}}{rank(x_i, l_k)}$	用于考察模型在不同标记上的准确率,同时考虑了标记的排序
One-error ^[17]	$one-error(f, \mathcal{U}) = \frac{1}{l} \sum_{i=1}^l \mathbb{I} \left[\left[\operatorname{argmax}_{l_k \in L} f_k(x_i) \right] \notin Y_i \right]$	用于考察样本的标记排序中排首位的标记不是相关标记的情况
Macro-F1 ^[18]	$macro-F1(f, \mathcal{U}) = \frac{1}{q} \sum_{j=1}^q \frac{2 \sum_{i=1}^l y_{ij} f_j(x_i)}{\sum_{i=1}^l y_{ij} + \sum_{i=1}^l f_j(x_i)}$	用于考察分类错误情况,其值为逐个标记计算后的平均F1
Micro-F1 ^[18]	$micro-F1(f, \mathcal{U}) = \frac{2 \sum_{i=1}^l f(x_i) \cap Y_i _1}{\sum_{i=1}^l Y_i _1 + \sum_{i=1}^l f(x_i) _1}$	用于考察分类错误情况,其值为标记矩阵中逐个元素计算后的平均F1
Macro-AUC ^[19]	$macro-AUC(f, \mathcal{U}) = \frac{1}{q} \sum_{j=1}^q \frac{ \{(a, b) \in Y_j' \times \bar{Y}_j' f_j(x_a) \geq f_j(x_b)\} }{ Y_j' \bar{Y}_j' }$	用于考察正确排序的分类错误情况,其值为每个标记的平均AUC
Micro-AUC ^[19]	$micro-AUC(f, \mathcal{U}) = \frac{ \{(a, b, i, j) (a, b) \in Y_j' \times \bar{Y}_j', f_i(x_a) \geq f_j(x_b)\} }{\left \sum_{i=1}^m Y_i \right \left \sum_{i=1}^m \bar{Y}_i \right }$	用于考察正确排序的分类错误情况,其值为标记矩阵中逐个元素计算后平均AUC

的表现,但是算法的性能可能受到标记排列顺序的影响。虽然问题转化方法在处理多标记学习问题时提供了灵活而有效的方式,但也可能面临一些挑战。例如,在面对大规模数据集或高维特征空间时,一些问题转化方法的计算复杂度比较高。

(2)从数据中直接学习的方法:这类方法试图在模型中直接考虑标记之间的关系和样本的多标记性质。因此,最简单的策略就是利用相似性度量方法直接对标记关系进行学习。例如,Huang等^[27]认为相似的标记应具有相似的类属性,因此利用余弦相似性方法计算标记的相似度矩阵来约束类属性的学习过程。此外,Lin等^[28]利用欧式距离计算两两标记在欧式空间下的相似性,从而达到衡量特征与标记之间的模糊关系的目的。此外,由于低秩学习的方法可以捕捉数据中的潜在结构^[29],因此可以利用张量分解和矩阵分解等技术对样本-标记关系进行建模,以便更好地理解标记之间的相关性和样本的关系。流形学习是一种用于处理高维数据的方法,其目标是通过学习数据的潜在流形结构,将高维数据映射到低维空间。流形学习基于一致性假设,即模型在输入空间中的相似点应该在输出空间中具有相似的预测^[30]。根据这一假设,相关研究主要从两个方面展开:其一,标记空间与特征空间具有相似的拓扑结构,利用计算到的特征流形可以重构标记空间信息^[31];其二,从多标记数据中学习目标对象的邻域信息,构建流形正则项对优化目标进行约束^[32]。还有一些其他常见的直接学习多标记任务的方法:主题模型^[33]、核方法^[34]、概率图模型^[35]等。这些直接学习的方法通常在捕捉多标记问题中的复杂关系上具有优势,而不需要显式地转换问题。

(3)优化损失函数的方法:在多标记学习中,优化损失函数的设计是关键的一步,因为它直接影响模型的性能。比如,Ranking loss用于考察模型的排序质量,即正确标记的排名在所有标记中的平均位置。早期的BoosTexter算法^[33]和Rank-SVM算法^[36]均是优化Ranking loss的代表性方法。另外,Xie等^[37]面向多源数据也采用了优化Ranking loss的方法进行一致性分类,同时还提出了一种优化Micro-AUC损失函数的方法^[38]。更多地,Wei等^[39]利用F1和Top- k precision来设计目标函数,学得分类器在处理缺失标记与未标注数据时均有很好的效果。尽管优化损失函数的方法在多标记学习中具有许多优势,但它们也可能存在一些缺点和挑战。例如,一些损失函数可能包含多个超参数,而选择适当的超参数值可能需要进行反复实验,这增加了调整模型的难度,并可能导致欠拟合。另一方面损失函数的效果可能对数据标记质量非常敏感。当存在标记错误或噪声时,一些损失函数可能会对其过度敏感,影响模型的性能。

(4)利用标记层次结构的方法:在多标记学习中,标

记之间可能存在一定的层次结构。这种层次结构可以体现标记之间的从属关系、上下级关系或其他形式的关联。利用标记层次结构的方法有助于更好地建模标记之间的关系,提高模型的性能。比如,Bi等^[40]改造了用于多标记学习的损失函数,使之适用于标记层次结构的学习,然后提出Bayes-Optimal的方法来对改进的损失函数进行优化。Wu等^[41]不仅利用图的方法保持学习过程中两两标记的相似度,还利用标记的层次结构构建约束项对标记关系作进一步制约。Wehrmann等^[42]提出新的全连接神经网络来实现具有层次结构的多标记学习。尽管利用标记层次结构的方法在多标记学习中具有许多优势,但也存在一些潜在的缺点和挑战,该方法的性能很大程度上取决于标记层次结构的质量,如果标记层次结构不准确或不完整,可能会导致模型学到错误的关系。

(5)利用标记局部关系的方法:在多标记学习中,利用标记局部关系的方法旨在捕捉标记之间的相对位置、依赖性或其他局部结构,以提高模型对多标记任务的性能。例如,Huang等^[43]最先提出标记局部关系的概念,标记局部关系指的是标记之间的局部相互关系,即某些标记之间可能更倾向于同时出现或相互排斥。通常使用图的结构来考虑标记之间的局部关系,例如通过邻接节点的信息来增强标记的预测。也有研究借鉴序列建模的思想,将标记序列作为输入,通过循环神经网络或类似的模型来学习标记之间的局部关系。Hou等^[31]沿用这一思想,利用局部线性嵌入的方法^[44]学习特征空间的拓扑结构,然后利用拓扑结构中包含的局部信息进行标记分布学习。此外,也有研究考虑在模型的输入中引入局部特征,这些特征与标记之间的局部关系相关。这可以通过对标记特定的局部特征进行建模^[45],或者引入标记相关的上下文信息来实现。近年来,有些工作考虑标记对之间的交互关系。可以通过训练模型来预测标记对的关联性,以便在预测过程中更好地利用标记之间的局部关系^[46]。

值得一提的是,Zhang等^[8]从标记关系利用的角度将多标记学习方法分成了三类,分别是一阶策略、二阶策略和高阶策略。一阶策略主要关注标记的独立性,即将每个标记视为独立的二元分类任务,在这种策略下,模型忽略了标记之间的相互关系。二阶策略引入标记之间的关联性,即考虑到标记对之间的依赖关系。高阶策略进一步考虑标记之间更高阶的关系,例如标记三元组或更大的标记组合。这样的方法可以通过使用更复杂的模型结构,例如高阶概率图模型或深度学习架构,来建模标记之间的更复杂的依赖性,不过高阶策略的计算复杂度也最高。

2 复杂数据背景下多标记学习面临的挑战

数据的不断增长导致学习对象的类别数越来越多,

建模也变得更加复杂。因此,近年来面向多义性对象建模的多标记学习技术受到重视。不仅如此,数据呈现的复杂特性又催生了多标记数据计算的其他问题。如前面提到的,数据的大规模、信息不完备、多维度、动态演化等特征制约着多标记学习模型的发展及其高效计算方法的设计和实现。针对这些特征,有必要对多标记数据展开研究并提供相应的处理方案。从应用角度来看,复杂数据在真实世界中普遍存在,对这样的数据进行挖掘是当前信息服务和科学发现的基石。比如,高维多标记数据通常指的是具有大量特征和多个标记的数据集,广泛存在于文本分析、多媒体内容理解和生物信息学等实际应用中^[47]。高维数据通常包含大量特征,这会导致计算复杂性的增加。同时随着维度的增加,模型需要更多的数据来泛化,否则可能会面临维度灾难的问题。再比如人工标注数据通常需要专业知识或者大量的人工劳动,这可能显著增加成本。尤其是对于多标记任务,每个样本可能需要被赋予多个标记,使得标注成本更为昂贵^[48]。从技术角度来看,数据的复杂性使得传统的数据处理和分析技术难以应对。作为一种适用于复杂数据处理的基础学习框架,多标记学习技术和应用发展是未来的必然趋势。开展面向复杂数据背景下的多标记学习更能反映真实场景,提高模型在实际应用中的适用性。在复杂数据背景下,样本包含了更丰富的信息,可以更全面地描述对象或现象。多标记学习允许模型从不同维度上理解数据,同时处理多个标记,有助于提取更丰富、更准确的特征表示。

基于上述分析,本章将重点对复杂数据背景下的多标记学习面临的挑战展开讨论。如图2所示,从标记、特征和样本三个层面,分别剖析多标记数据特性带来的挑战及其内在机理,为后续解决方案的设计提供清晰的问题导向。

2.1 标记层面

对于复杂数据背景下的多标记数据,一个重要的体现是标记的大规模性^[49]。众所周知,多标记数据中类别繁多且语义相对复杂。随着数据不断产生和累积,待建模类别不断扩张,标记体量呈现爆发式增长态势。因此从一个庞大的标记空间中为目标对象找到一个相关的标记子集需要付出大量的计算成本,这给多标记学习建模带来了巨大的挑战。其次,学习对象丰富的语义信息使得其监督信息难以全部获取,现实中多标记数据的标记信息通常是不完整的。其中,最典型情况是只有少量样本有完整标注数据,其他大量样本为未标注数据^[50]。举例来说,在图像标注任务中,海量图像可以在网络上获得,然而由于标注成本过高,只有少部分的数据被人工标注。另一种典型情况是多标记数据中存在标记缺失。以社交媒体内容分类为例,标注工作者可能只关注用户发布的一些热门主题或常见关键词,而忽略了一些模糊、不常见或特殊领域的内容。更极端地,某些标记信息完全缺失,使得与这些标记相关的训练样本无法获得。具体地,对于某些标记而言,训练集和测试集之间没有任何交集,利用训练数据生成的模型需要对未见过的对象类中的测试样本进行预测^[51]。以往的多标记学习框架遵循的基本假设是:与样本所有相关的类别标记都有提供用于训练。那么基于这样的学习框架研究的多标记学习方法,已然不适宜处理这些复杂情形。

多标记数据在标记层面计算的复杂性还体现在标记空间具有动态演化的特点^[52]。数据流是现实生活中数据产生的一种重要形态。事实上,多标记数据中的标记不是或不能事先给定,而是随时间动态变化^[53]。比如,当新的样本到来时,可能出现原先标记空间中没有的新标记^[54]。面向流标记数据分类,需要构建在线学习模型对新来的标记做出判断和预测。

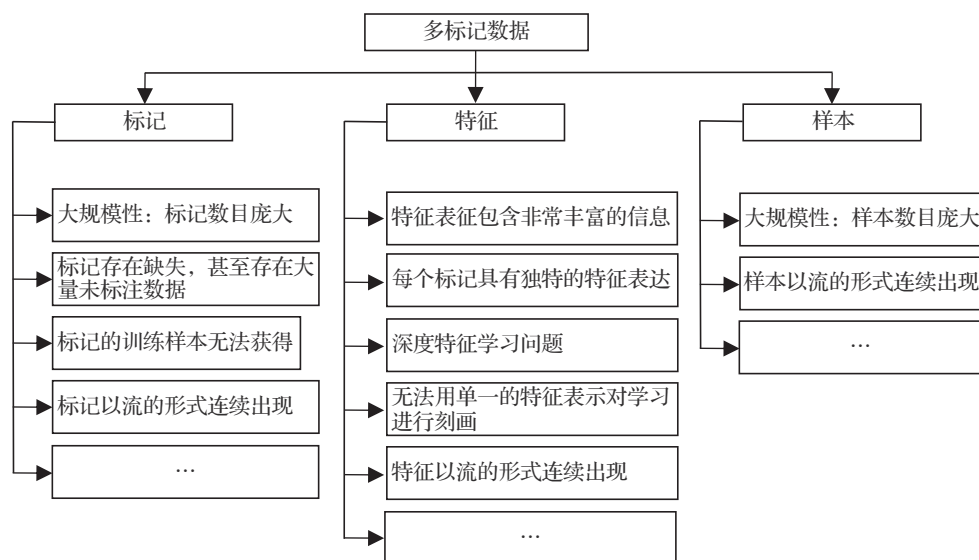


图2 复杂数据背景下多标记学习面临的挑战

Fig.2 Challenges of multi-label learning under complex data environments

2.2 特征层面

从特征层面考察复杂数据背景下的多标记数据,成因与学习对象丰富的语义信息密不可分。正是由于学习对象这一特质,表示该对象的特征信息通常也非常丰富,丰富信息带来利好的同时也使得多标记数据学习问题变得棘手。

从特征信息的表示说起,采集到的多标记数据多以没有预定义的非结构化形式存在,如图像、音频和文本等。在解决实际任务时,除了对非结构化数据直接进行处理,另一种常见的表示方法是从非结构化数据中进行特征提取来获得特征信息结构化表示。鉴于特征信息非常丰富,结构化的特征表示空间往往具有高维性,从而引发“维数灾难”问题^[55]。以“雅虎”数据集为例,该数据集从文本数据中收集而来,并用多个类别标记来表示文本的不同主题。因此,数据子集的标记个数在21到40之间变化,而提取用来表示文本的特征向量最低有21 920维,最高达52 350维。鉴于此,从结构化或非结构化多标记数据中获取对标记有辨识力的特征表达,甚至为每个标记构造一个其自身独特的特征表达,成为特征层面复杂数据背景下多标记学习的主要体现。丰富的特征信息也会造成单一的特征表示无法很好地描述学习对象。在多标记学习中,每个学习对象通常由一个示例表示。然而,这种表示方法过于简单,在表示阶段会丢失有用的特征信息。事实上,每个学习对象应该是由多个示例表示。例如,在图像标注中,若用某种技术将图像分割成多个区域,每个区域用一个示例表示,那么利用多个示例便可以更好地反映该图像复杂繁多的语义信息^[56]。因此,特征层面计算复杂性的另一个体现是多标记数据的多示例表示。

最后,在多标记学习中,特征空间也具有动态演化的特点。与动态标记背景呈现的特点类似,学习问题的特征空间也不是事先完全获知的,通常是以流的形式连续不断地到来,即随着时间推移,可能出现新的描述学习对象的重要特征^[57]。考虑到特征空间的动态性和未知性问题,在多标记学习过程中有必要建立新来特征的在线处理机制。

2.3 样本层面

从样本层面上看,复杂数据背景下的多标记数据计算同样存在其固有点,这在描述标记和特征计算复杂性时已有所体现。比如,样本数目可能非常庞大,具有大规模的特性。正因如此,在学习过程中难以保证多标记数据计算的高效性。从另一个角度来看,庞大的数据量在强大算力的支撑下有望实现学习性能的显著提升。所以不论是追求计算效率还是追求学习性能,样本的大规模性都给多标记学习建模提出了更为苛刻的要求。再比如不只是标记和特征具有动态性,样本同样存在动态出现这一特点。在新样本增量流入学习系统时,学习模型需要有能力对新样本的类别标记进行预测^[52]。由此可见,有必要展开相关的算法研究,使多标记学习算法适用于各种复杂形式数据的处理和分析。然而,研究人员往往会将样本呈现出的复杂性特征延伸到标记层面和特征层面进行考虑,因为多标记数据的计算问题本质上还是为了建立一个从特征空间到标记空间的映射,用以实现多标记学习。

正如前文所提及的,样本的大规模性会激发从特征层面学习原始数据的深度特征表达,继而完成多标记分类;新样本的流入会引起之前未知的新标记出现,新样本的处理可看作流标记学习的必要步骤。因此,在接下来的章节中,主要针对复杂数据背景下在标记和特征层面进行多标记学习的相关算法研究进展进行综述。

3 复杂数据背景下的多标记学习算法

本文将复杂数据背景下多标记学习相关的算法划分为两大类,即面向标记复杂性计算的方法和面向特征复杂性计算的方法,如图3所示。在标记层面,可以进一步划分为超高维标记分类学习算法、弱监督多标记学习算法、零样本多标记学习算法和流标记学习算法。而在特征层面,可以进一步划分为多标记特征选择算法、类属性学习算法、基于深度模型的特征学习、多示例多标记学习和流特征在线多标记学习。

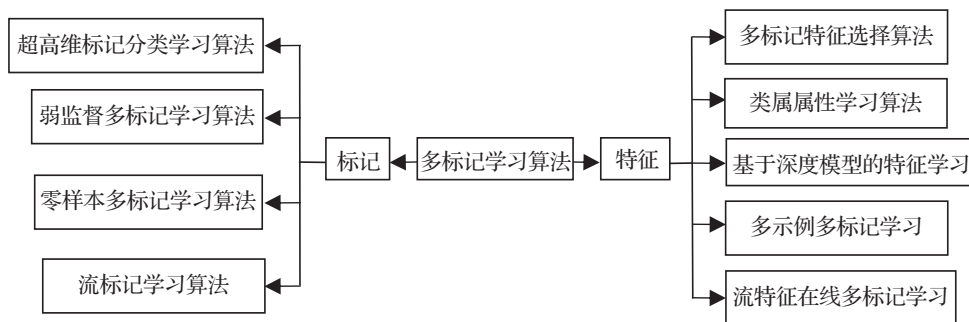


图3 复杂数据背景下多标记学习算法分类

Fig.3 Classification of multi-label learning algorithms under complex data environments

绍不同的多标记学习算法研究现状。

3.1 面向标记复杂性计算的方法研究

3.1.1 超高维标记分类学习算法

超高维标记分类学习涉及处理具有大量标记的数据集,每个样本可能同时属于多个标记^[58]。这种情况常见于现实世界中的复杂任务,例如图像多标记分类、文本多标记分类、生物信息学等领域。常用的超高维标记分类学习方法主要有基于嵌入的方法、基于树的方法和一阶多标记学习方法。这些方法在不同层面缓解了高维标记空间带来的挑战,但也存在各自的局限性,下面对其进行系统梳理与分析。

(1)基于嵌入的方法:假设高维标记空间具有低秩结构,先对其进行低维嵌入表示,然后构建特征空间到该低维嵌入的线性映射,最后生成所有标记的输出结果^[59]。而这一假设却与超高维标记数据中的长尾分布规律相违背,导致学习到的低维嵌入难以恢复原有的标记信息。然而,该方法的核心假设是标记空间低秩性,这在实际数据中往往不成立。此外,此类方法通常忽略了原始特征空间与标记空间之间的结构关系,生成的低维嵌入无法保证原始标记与特征空间数据分布的一致性^[60]。

(2)基于树的方法:该类方法从训练数据中学习归纳出一个树型层次结构,然后利用信息增益或者基尼系数等指标进行剪枝^[61]。与基于嵌入的方法不同,树方法更强调结构划分与快速索引,常用于对预测效率要求高的多标记分类任务。具体来说,首先对根节点进行初始化,然后根据特定的节点划分方案,如优化 F 值和 $nDCG$ 排序损失逐一划分每个非叶子节点的样本空间,最后为每个叶子节点分别构建一个基分类器,每个基分类器只负责对应的叶子节点所包含的少量标记^[62]。此外,还可以通过标记嵌入进行层次聚类构建标记树,从而利用分类器在路径上搜索得到样本的标记集合^[63]。因此,基于树的方法通常具有很高的计算效率,训练时间可以达到数据规模的次线性甚至更低的复杂度。然而,构建一棵全局最优树往往代价极高,且高度依赖超参数设置,如树深度、划分标准、剪枝策略等。此外,当标记分布不均、标记之间缺乏层次关系时,树结构划分

容易失衡,影响整体性能。

(3)一阶多标记学习方法:一阶方法将多标记学习问题转化为一组独立的二分类任务,该类学习策略由于其简单直接,使其在超高维标记分类学习中受到重视,但在标记信息利用和计算效率等方面存在不足。为此,一些研究人员尝试提出改进方法,使之能高效处理超高维标记数据。然而,由于本质上各标记独立建模,这类方法难以捕捉标记间的结构性关联,可能导致整体预测结果的不一致性^[64]。

虽然上述提及的方法均在超高维标记分类学习任务中取得了不错的性能,但未涉及对超高维特征引发的不确定性进行建模。因此研究人员开始采用标记选择的方法来减少标记空间维度,用以构建高效的超高维标记分类机制。通过探索尾部标记对常见超高维标记分类评价指标的影响来缩小用于建模的标记规模,或者采用随机采样策略搜寻一个重要标记子集来近似表示原始标记信息,以此进行标记选择来分步设计特征空间到超高维标记空间的映射。

上述方法在应对超高维标记空间的挑战中各具优势。嵌入方法通过空间压缩提升效率,树结构方法强调结构组织与推理速度,一阶方法凭借简单性快速迭代,标记选择方法则在维度削减上独具成效。但从整体来看,当前研究中维度约减与性能提升之间存在天然矛盾,尚需进一步优化平衡策略。此外,缺乏对标记和标记、特征和标记之间多层次结构性关系的全面建模。表2进一步总结归纳了超高维标记分类典型方法的优缺点。

3.1.2 弱监督多标记学习算法

多标记学习中每个样本可能属于多个标记中的任意组合,使得标记数据的获取常面临高成本、高噪声与不完整等现实限制,导致完整、准确的训练数据难以获得^[74]。为了解决上述困难,弱监督多标记学习应运而生。弱监督多标记学习是多标记学习的一种形式,其中训练数据的标记信息相对不完全。弱监督多标记学习可以根据监督信息化为弱标记学习、半监督多标记学习和偏多标记学习。其中,弱监督多标记学习中标记信息是不完全准确、噪声较多或只提供了部分信息。半监督

表2 超高维标记分类典型方法总结

Table 2 Summary of representative methods for ultra-high dimensional multi-label learning

算法类型	代表算法	主要机制	优势	局限性	适用场景
基于嵌入的方法	SLEEC ^[65] REML ^[66]	将高维标记空间映射为低维嵌入空间进行学习	可解释性强,计算效率高	对低秩假设依赖强;忽略特征-标记空间关系	高维、稀疏的极端多标记分类任务
基于树的方法	FastXML ^[67] GBDT-SPARSE ^[68]	构建标记层次结构或树状结构后建模	具有较高的预测准确率和效率	树结构构建复杂;超参多,对非层次标记不友好	大规模、高维、含层次结构的标记
一阶多标记学习方法	PfastreXML ^[69] PPDspase ^[70]	每个标记单独分类,或引入并行/稀疏优化	并行计算,考虑标记之间的层次关系和依赖关系	忽略标记依赖,鲁棒性低	对预测速度要求高的学习场景
标记选择方法	ADATTL ^[71] CSSAG ^[72] moplms ^[73]	筛选重要标记子集以降低维度	降低建模难度	可能丢失部分信息	标记空间高维且含有冗余和噪声标记,可解释性高

多标记学习中,标记信息可以是完全准确的,但只有一部分样本被标记,其余样本没有标记信息。而偏多标记学习是弱监督多标记学习的一个特例,每个实例相关的标记集合中只有一部分是真正相关的,而其他部分可能是由于噪声或其他原因而被错误地关联,因此偏多标记学习的目标是识别出每个实例的真实相关标记集合。

弱标记学习的核心思想是通过挖掘数据中隐含的关联信息,弥补缺失或错误的标记,从而提升模型的预测能力。例如,Jing等^[75]提出了一种半监督低秩映射学习算法SLRM,该算法通过学习一个低秩映射矩阵挖掘特征关系,然后利用已标记数据进行训练并学习一个初步的分类器,再利用未标记数据和初步分类器进行迭代优化,更新分类器和映射矩阵,最后利用更新后的分类器和映射矩阵对未标记数据进行预测。此外,考虑到标记缺失可能导致类别分布不平衡的问题,因此可以采用处理类别不平衡数据的方法间接应对缺失标记的问题。例如,Dong等^[76]充分利用了集成学习的优势提出了半监督弱标记学习算法SSWL。该算法通过构建线性分类器来分别学习正类和负类标记信息,同时通过最大化两个分类器对于预测结果的一致性,实现了对类别不平衡和标记缺失的联合处理。

半监督学习主要有两种策略,分别是归纳式学习方法和直推学习方法。归纳式学习方法通过把未标注数据当作训练数据进行学习建模,然后再对测试数据进行预测。例如,iMLCU^[77]是第一个用于半监督多标记学习的归纳式方法,该算法通过考虑两两标记之间的相关性并对未标记数据施加最大间隔正则化。直推学习则通过多个模型之间的协同训练,充分利用特征集的互补性。例如,Kong等^[78]提出了一种利用标记传播算法进行半监督多标记学习的算法TRAM,该方法基于图的模型来构建已标记数据和未标记数据之间的拓扑结构,然后使用这个拓扑结构来传播标记。在传播标记的过程中,不仅考虑了数据之间的相似性,还考虑了标记之间的相似性,从而提高了传播标记的准确性。归纳式方法

适用于泛化任务,直推式更强调局部传播效果。但也各自存在弱点,例如归纳式方法对有标记样本依赖较强,数据不足时易过拟合;直推式方法计算复杂度较高,传播过程容易被噪声放大。

偏多标记学习处理的问题是样本标记集合中存在部分错误标注,需挖掘出真正相关的标记子集。按照策略不同可以分为基于标记处理的方法、基于模型调整的方法和基于集成的方法。基于标记处理的方法通过推断缺失标记、制定规则或利用图模型直接处理部分标记数据,以预测完整标记集合。例如,CPLL算法通过期望最大化模型迭代估计缺失标记的概率和更新模型参数,从而实现预测每个实例的完整标记集合^[79]。PLAIN利用图模型来表示实例和标记之间的关联,通过图算法(如最短路径、社区检测等)来推断标记集合^[80]。这类方法在噪声标记场景中表现良好。然而,其主要难点在于真实标记的不可见性常导致推断误差积累,如何在鲁棒性与泛化能力间取得平衡仍是重要挑战。

弱监督多标记学习通过不同程度整合未标注、缺失、错误信息,扩展了传统多标记模型的适用范围,在实际应用中具有广泛价值。但仍然存在一些问题:一方面图建模方法计算复杂度较高使得在大规模数据集上扩展较难;另一方面模型稳定性与泛化性不足,尤其是在弱标记或偏多标记问题中,错误传播容易导致预测出现较大偏差。表3进一步总结归纳了弱监督多标记学习典型方法的优缺点。

3.1.3 零样本多标记学习算法

多标记零样本学习是一种对训练过程中未定义的未知样本进行预测的特殊学习算法^[87]。尽管当前大量的研究都是基于人工标注的训练数据集进行训练,但一些稀缺类别的样本极难获取,使得学习算法不能够依赖大规模训练样本进行学习^[88]。因此,研究在目标域无标注数据集上进行有效的零样本学习具有重要意义。此外,基于此学习范式演化的单样本学习和小样本学习都已经在真实场景中得到了有效应用^[89]。零样本学习需

表3 弱监督多标记学习典型方法总结

Table 3 Summary of representative methods for weakly supervised multi-label learning

算法类型	代表算法	主要机制	优势	局限性	适用场景
弱标记学习	GLOCAL ^[30]				
	SLRM ^[75]	低秩分解、图结构传播等	可处理缺失与噪声	对低秩假设依赖强;忽略特	标记缺失率高、噪声多、标记存
	ML-LRC ^[81] MMIB ^[82]	学习标记相关性来恢复	标记;挖掘隐含结构	征-标记空间关系	在隐含关联或低秩结构的场景
半监督学习 (归纳式)	iMLCU ^[77]	利用少量有标记数据引导	具有较高的泛化性	标记信息稀少时泛化弱;对	标注成本高、未标记数据充足、
	COINS ^[83]	模型训练未标记样本	能和效率	模型初始训练要求高	数据分布复杂场景
	DRML ^[84]				
半监督学习 (直推式)	TRAM ^[78]	图建模和标记传播,利用	捕捉标记/样本关系,	图构建成本高,传播机制易	样本结构明确、标记协同性强、
	CMLP ^[85]	结构一致性引导学习	提升协同性能	受噪声干扰	规模中等的场景
	MLRMG ^[86]				
偏多标记学习	CPLL ^[79]	图推断、概率建模、标记集	适合处理部分错误	真实标记推断不易,错误传	众包标注、自动标注输出及复
	PLAIN ^[80]	合还原	关联标记数据	播风险大	杂易混淆标记场景

要基于已知标注数据学习并预测未知样本的标记,因此需要结合已有知识(如词向量和知识图谱)对未见样本实现正确分类。参照零样本单标记学习算法的分类方式,零样本多标记学习的主流学习算法可以分为:基于直接映射的方法、模型参数传递方法和联合潜在空间学习方法。

(1)基于直接映射的方法:该类方法致力于学习从特征空间到语义空间的映射函数,并在语义空间中完成标记预测。与单标记零样本学习不同,多标记任务需同时考虑多个标记及其组合属性,从而面临标记集不一致、标记数量变化大等挑战。为了解决这个问题,现有的研究使用所有标记的组合属性(如词向量)的平均值用于集体映射,从而达到零样本多标记学习的目的。然而,这种组合语义表示方式也会造成语义丢失。例如, Fu等^[90]最早提出零样本多标记学习算法TraMP,该方法利用直推式算法来捕获标记的集体语义表示,然后分别使用最近邻搜索和辅助标记信息来寻找匹配最优的标记,从而实现零样本多标记学习。但是,该方法没有解决面临的标记组合的指数爆炸问题。此外,各标记和示例间的相关关系存在优先级别,因此根据此优先级信息来约束目标函数,可以提升模型的局部感知能力。

(2)模型参数传递方法:该类方法通过在已知标注数据上学习模型参数,然后为未知样本建立模型,代表方法COSTA^[91]为每个标记训练单独分类器,并利用标记共现性传播模型参数。然而,该方法忽略了标记间的语义结构,共现性在标记稀疏或噪声标记下的稳定性较差。模型参数传递方法关注于利用已知类别的学习参数,迁移生成未知标记的预测模型。这种方法在样本稀缺情况下表现出较强的迁移性,但往往忽略了标记语义之间的高阶关系,且依赖共现性建模时易受到噪声影响。

(3)联合潜在空间学习方法:该方法通过在视觉空间和语义空间联合学习到一个潜在空间,并在此嵌入空间上缩小语义之间的差距,从而实现零样本多标记学习,因此广泛应用于多媒体信息检索领域^[92]。例如, Wang等^[93]提出了一种联合视觉和语义嵌入空间的学习算法BiDiLEL,促进语义鸿沟和知识迁移,从而实现零样本多标记学习。但是此类方法在真实场景中的应用还存在一定差距,实时性上还需要进一步改进。联合潜

在空间方法致力于在语义与视觉空间之间构建统一嵌入空间,从而提升语义对齐能力。此类方法在多模态环境下有良好适应性,但模型结构复杂、训练开销大。

直接映射方法简洁高效,但难以表达标记组合结构。模型参数迁移可快速适配新标记,但存在估计偏差。联合潜在空间方法表达能力强,适合多模态场景,但训练代价高。表4进一步归纳总结了零样本多标记学习典型方法的优缺点。

3.1.4 流标记学习算法

该学习方法研究的是多标记数据在动态背景下如何构建学习模型,使之在时间变量的影响下仍然有好的性能表现,从而构建标记流的在线处理机制^[98]。该任务具有显著的实际意义,如舆情分析中新兴话题的自动发现、推荐系统中新兴趣标记的动态更新等。早期研究通常假设样本空间保持不变,而专注于设计优化方案以应对标记空间的动态变化。例如, Lin等^[99]提出了一种对新标记到达后的特征空间排序的优化算法MLFSL。MLFSL定义了最终的特征排序的秩列表和每个标记的特征排序的秩列表,最终的优化目标就是使得两个秩列表的总体加权偏差最小化,从而实现特征空间的排序。尽管现有的流标记学习算法在建模新标记时取得了一定效果,但在样本流持续输入的现实应用中,它们无法处理训练样本随时间推移不断增加的挑战。为了使学习算法不仅可以在已知标记的样本上取得很好性能,而且还可以在未标注的新标记充分识别分类,一些有代表性的成果相继被提出。例如, Pham等^[100]提出了一种基于多标记多示例学习框架MIMLNC,用于流标记数据处理。MIMLNC通过动态规划进行似然估计,使用迭代优化模型最终发现测试集中隐含的新语义,并准确预测测试集样本中是否包含这些新的语义。但是由于该算法每次运行都需对所有训练数据进行遍历,所需的时间消耗与标记量指数相关,因此很难使用其进行大规模数据处理。为此, DMNL算法^[101]进行了改进,使用一种无监督的方法对实例进行聚类,然后对每个聚类进行标记预测。在预测标记时,它不仅考虑了每个实例的特征,还考虑了该实例所属的聚类。由于这种方法考虑了实例之间的相似性和差异性,因此它可以发现那些在现有标记集中没有出现过的新的标记。但是,该方法

表4 零样本多标记学习典型方法总结

Table 4 Summary of representative methods for zero-shot multi-label learning

算法类型	代表算法	主要机制	优势	局限性	适用场景
直接映射	TraMP ^[90] Fast0Tag ^[94] MIVSE ^[95]	从特征空间直接映射到语义空间,进行标记预测	实现简单,模型轻量,适配性强	标记组合存在语义丢失;难处理标记数量不一致	标记数量少,实时性要求高及特征标记线性关联
模型参数传递	COSTA ^[91] EXEM ^[96]	利用已有类的模型参数推断未知标记分类器	利用已学知识迁移,模型更新灵活	忽略标记语义信息和标记间依赖性;对低频标记敏感	样本稀缺或资源受限场景
联合潜在空间学习	BiDiLEL ^[93] ConSE ^[97]	联合视觉与语义空间构建通用潜在表示	捕捉标记样本关系,提升协同性能	图构建成本高,传播机制易受噪声干扰	语义鸿沟显著,标记语义关联丰富,数据规模适中

只针对静态测试数据集进行新标记预测。为此, Zhu等^[102]设计了一种流标记学习算法 MuENL, 通过同时对特征和标记预测值构造新的检测器, 并且为发现的新标记设计鲁棒性强的更新模型。

总体而言, 流标记学习通过设计对新标记敏感的模型更新机制, 使多标记分类具备更强的实时性和泛化能力。但大多数方法在标记数量增加时复杂度迅速上升, 难以适应大规模标记集。此外, 现有流标记学习算法对标记相关性建模不足, 多数方法仅依赖局部相似性或聚类结果, 未能充分挖掘标记之间的潜在语义关联, 导致对新标记预测的鲁棒性有限。表5进一步总结归纳了流标记学习典型方法的优缺点。

3.2 面向特征复杂性计算的方法研究

3.2.1 多标记特征选择算法

特征选择技术可以从大量数据中识别和选择最相关的特征, 提高模型的效率和准确性。多标记学习任务中, 样本通常关联多个标记, 特征维度与标记空间维度均可能极高, 特征选择面临更大的挑战。此类任务不仅需要选择与预测目标强相关的特征, 还需考虑标记之间复杂的相关结构^[104]。因此, 如何融合标记依赖、特征选择和降维过程成为关键问题。

(1) 基于稀疏学习的方法: 该类方法旨在通过考虑特征的稀疏性来降低数据的维度, 有助于处理高维度的多标记数据, 减少冗余信息, 提高模型的效率和泛化能力。常用的稀疏正则项主要有 l_1 正则化^[105]和 $l_{2,1}$ 正则化^[106]。其中, l_1 正则化引入稀疏性, 通过在损失函数中添加 l_1 范数项来惩罚不重要的特征。这导致一些特征的权重变为零, 从而实现了特征选择和降维的效果。 $l_{2,1}$ 正则化通过添加权重的范数平方的一部分作为损失函数的一部分, 以防止模型的权重过大。因此, 将重要度信息、标记关系或者类属属性等多标记数据的特性结合, 从而可以实现特征选择的目的。此外, $l_{2,0}$ 指的是 l_2 范数和 l_0 范数的混合形式, 即对权重向量的 l_2 范数进行惩罚, 但只保留其中非零元素^[107]。这种混合范数的优化问题可能比 l_2 范数更为复杂, 因为它涉及到非凸优化

问题, 因此对于大多数机器学习任务, 它并没有被广泛采用。 l_2 范数不能用于生成严格的稀疏解, 然而有研究证实 l_2 范数的约束同样有助于重要特征的辨识。这类方法的优点是可以联合建模特征选择与分类器学习过程, 尤其适合处理“样本少、特征多”的灾难性数据。但在优化时可能陷入局部最优, 对正则系数敏感, 部分模型训练复杂度较高。

(2) 基于信息理论的方法: 该类方法通过利用信息度量的方式来评估特征的重要性, 将原始特征分为强相关特征、弱相关非冗余特征、弱相关冗余特征和无关特征, 而搜寻强相关特征和弱相关非冗余特征是学习的目标。因此, 早期的研究主要是基于最大相关性最小冗余性的学习框架来实现多标记特征选择。例如, PMU^[108]使用了一种基于互信息的多标记特征选择准则, 特点是它通过最大化所选特征与标记之间的依赖关系来选择有效的特征子集, 这是第一次提出一个特征过滤准则, 在评估给定特征的依赖性时考虑到标记的相互作用, 而不诉诸于问题转换。由于现有的基于信息理论的方法采用过滤式的特征选择方式, 特征子集的生成过程往往需要耗费大量的时间, 且结果容易陷入局部最优。为处理这一问题, 相关研究提出了多种方法进行改进。例如, 通过对信息度量方法和搜寻策略进行优化改进, 从而避免冗余计算过程^[109]。另外, 一些研究还致力于挖掘多标记数据中的关联关系, 从而可以确保所选择的特征子集具有全局最优性质, 并且为更深入理解多标记数据提供了更多技术实现的可能性^[110]。

此外, 对于多标记特征选择的研究还有许多探索。比如, Zhang等^[111]提出一种采用两种不同的映射方式的方法, 旨在最大程度保持特征与标记空间的相关性。Zhang等^[112]提出了一种经典的嵌入式学习方法, 通过将汉明损失和排序损失作为优化目标, 进而使用遗传算法来搜寻最优特征子集。Yan等^[113]通过标记关系构建间隔损失函数, 以实现多标记特征选择。Huang等^[114]通过改变样本和标记之间的相似度计算策略, 利用 Laplacian score 评估每个特征的重要性。综上所述, 基于稀疏学习的方法更侧重于在高维空间中构造稀疏结构, 提升模

表5 流标记学习典型方法总结

Table 5 Summary of representative methods for streaming label learning

算法类型	代表算法	主要机制	优势	局限性	适用场景
仅标记空间变化	MLFSL ^[99] FSSL ^[103]	通过特征排序、类间判别性学习等方法适应新标记的加入, 调整模型参数	动态调整特征选择策略, 提升对新增标记的适应能力	忽视样本空间变化, 模型更新不及时, 难以应对流式数据	样本分布不变, 标记增量扩展, 特征关联明确及资源受限
样本和标记空间联动变化	MIMLNC ^[100] DMNL ^[101]	利用动态似然估计、聚类优化等方式联合建模标记和样本空间的变化	能挖掘未知标记, 提高模型对潜在标记的发现能力, 适用性强	计算复杂度高, 实时性有限, 部分方法依赖静态数据训练	样本分布不变, 标记增量扩展, 特征关联明确及资源受限
其他方法	MuENL ^[102]	构建新标记检测机制与增量更新模块, 实现在线学习与模型扩展	支持新标记的及时检测与鲁棒学习, 适用于流数据环境	对早期误报较敏感, 初始阶段易受干扰	新标记频繁涌现, 资源受限需轻量化更新

型的泛化能力。但其优化问题通常难以获得全局最优解,且正则项权重选择敏感。而基于信息理论的方法强调通过信息增益、互信息等度量方式捕捉特征与标记之间的统计依赖关系,但也带来了计算复杂、特征与分类过程解耦等问题。表6进一步总结归纳了多标记特征选择典型方法的优缺点。

3.2.2 类属属性学习

类属属性学习假设每个标记具有其内在独有的特征,而这些特征对相应标记的预测能力很强。该类方法旨在为每个标记构建最具区分度的特征子空间,从而提高多标记分类模型的预测准确性和泛化能力^[125]。已有一些工作尝试采用聚类、稀疏学习、多目标优化的方法来进行类属属性学习。Zhang等^[19]最早引入了类属属性的概念,并将其应用于多标记学习,通过采用聚类方法为每个标记生成类属属性。具体而言,该方法首先对每个标记的正类样本和负类样本进行了聚类分析,然后利用聚类中心得到了每个标记的类属属性空间。最终,通过生成的类属属性空间对相应的标记进行预测。后续,FRS-LIFT^[126]和LSDM^[127]方法都在此基础上借鉴了这一思想实现多标记学习方法。此外,Huang等^[128]提出了一种采用稀疏学习方式的LLS算法,能够为不同的标记找到具有鉴别能力的类属属性集合。Kashef等^[129]采用了多目标优化策略,并将其引入到类属属性学习中,从而通过寻找Pareto前沿来解决多目标优化问题并得到包含了不同标记的类属属性的最优解集。目前,更为常见的研究方向是将类属属性学习融入算法设计的整体框架,以解决其他更为复杂的多标记学习问题。例如,Zhang等^[130]提出一种结合标记相关性的类属属性多标记分类方法MLFC,探索稀疏约束的类属属性选择,并尝试将标记相关性用优化过程生成的额外特征进行表示。因此,类属属性学习通过细粒度地刻画标记与特征之间的独特关系,有效提升了多标记学习的表达能力。然而,其也面临一些挑战,例如如何高效处理高维稀疏数据以及如何应对标记间属性重叠等问题。

3.2.3 基于深度模型的特征学习

随着深度学习技术的发展,其强大的特征表达能力和非线性建模能力使其在多标记学习任务中表现出显著优势。相比传统模型,深度神经网络能够更充分地捕捉样本中的高层语义信息,同时也为建模标记之间复杂依赖关系提供了强大的结构表达能力^[131]。在多标记学习中,基于深度模型的特征学习对于有效地捕获图像或文本中多个标记之间的关联关系至关重要。Zhang等^[132]最早提出了利用神经网络结构进行多标记学习的算法BPMLL,该方法基于后向传播神经网络来捕获多标记数据的特征,并且显式地利用标记之间的依赖关系实现多标记学习。后来,相关研究表明,BPMLL在文本数据集的表现并不那么令人满意,为此Nam等^[133]提出了一种相对简单的神经网络架构NN-MLL,利用该方法可以在大规模多标记文本数据集上达到最优的结构。另外,现有的嵌入方法虽然可以有效地获取标记之间的依赖关系和降低计算成本,但是这些嵌入都是浅层的,无法发现标记之间的高阶关系。为了缓解这一挑战,Yeh等^[134]基于深度神经网络框架提出了C2AE算法,该方法改进了深度典型相关分析方法和自编码器对标记和特征做联合嵌入,然后通过引入标记相关的损失函数进行标记的预测输出。随后,许多基于C2AE改进的方法相继提出,如DCSPE^[135]和DBPC^[136]等算法。

为了探索潜在的标记结构信息,越来越多的神经网络架构被用来进行多标记学习,例如Wang等^[137]提出了一种利用循环神经网络进行高阶标记依赖性学习的方法CNN-RNN,该方法通过学习图像和标记之间的相关性进行端到端的训练。近来,图卷积神经网络也被应用于多标记学习中。例如,Chen等^[138]提出了在标记上构建有向图,然后利用图卷积神经网络对标记之间的相关性进行建模,最后将标记的嵌入应用到分类模型的设计。另外,作为深度神经网络模型的替代方法,Yang等^[139]提出了利用基于树模型的集成深度框架Deep Forest进行多标记分类。该方法通过优化不同的性能指标和

表6 多标记特征选择典型方法总结

Table 6 Summary of representative methods for multi-label feature selection

算法类型	代表算法	主要机制	优势	局限性	适用场景
稀疏学习方法	MDFS ^[115]				
	GLFS ^[116]	通过引入 $l_1/l_{2.1}$ 正则约束、特征-标记建模等方式实现稀疏特征选择	处理高维稀疏数据能力强,支持端到端优化	易陷入局部最优,对正则系数敏感,非凸优化求解困难	高维稀疏数据,特征冗余噪声多,可解释性要求高及端到端学习场景
	BDLS ^[117]				
	JFSC ^[118]				
LLSF ^[127]					
信息理论方法	PMU ^[108]				
	MDMR ^[119]				
	MIFS ^[120]	基于互信息、信息增益、条件互信息等度量评估特征对多个标记的贡献	理论清晰,适用于标记结构复杂任务,便于分析与解释	特征选择与分类解耦,计算量大,易陷入局部最优	多标记关联复杂,特征-标记非线性,冗余特征多,可解释性要求高及小样本/未知分布场景
	IGME ^[121]				
	LRFS ^[122]				
WFRFS ^[123]					
STFS ^[124]					

减少过拟合来提升模型的性能,并且大量实现表明 Deep Forest 可以取得更优异的性能。

基于深度学习的方法在多标记学习中表现出强大的特征提取与标记依赖建模能力。这些方法大多通过端到端结构同时优化特征表示与标记输出,显著提升性能。但存在的共性问题网络结构设计复杂、训练开销大,并且对大规模标记集或缺少标记先验知识的数据,效果不稳定。

3.2.4 多示例多标记学习

由于真实世界中多义性对象的存在,导致很难利用监督学习框架进行有效建模。同时,如果建模过程中只利用其中一个示例来描述学习对象,往往会导致有效信息的丢失,从而使得模型的学习性能下降。为了解决上述问题,Zhou 等^[140]最先提出了多示例多标记学习的概念。多示例多标记学习模型由于其强大的表示能力,受到了研究者的广泛关注,各种学习算法不断被提出,相关算法可以分为:基于退化方法和基于正则化方法的多示例多标记学习算法。

基于退化方法的多示例多标记学习算法将原始的多示例多标记学习问题转化为多标记学习或者多示例学习问题,从而利用相应的学习算法实现问题的求解。Zhang^[141]最先将 k 近邻分类方法推广到多示例多标记学习,并提出了 MIML-kNN 算法。MIML-kNN 算法通过利用近邻传递信息以及对不可见示例进行多标记预测。但是,该方法在训练过程中可能会丢失训练示例中的有用编码信息,从而导致学习性能的退化。退化方法思路简洁,便于复用已有模型,其天然缺陷在于忽略原始多示例多标记学习问题中的联合结构建模需求,特别是示例与标记之间复杂而隐含的依赖关系。

虽然可以通过将 MIML 学习问题转化为多示例学习或者多标记学习问题进行解决,但是由于采用退化转换的策略会导致样本中重要隐含信息的丢失,从而使得标记之间或者示例与标记间的关联性减弱^[142]。为此,很多学者相继提出了将标记信息融入正则化的方式。Zha 等^[143]提出了一种全新的 MIML 学习算法 MLMIL,通过利用正则化方法,并使用样本示例和概念标记之间的联合,充分考虑了示例和标记之间的潜在的联合关系,并且在样本示例集合上直接进行建模。但是这种对样本包与标记、包内示例与标记之间相关关系的建模,所构

造的模型复杂度高,对大规模数据场景可能不太适应。

通过对当前主流多示例多标记学习算法的分析可知,基于退化方法的算法实现简洁、可直接复用传统模型,适合小规模或特征清晰的任务场景。但其通用性受限于退化过程的适配能力,且在转化过程中容易丢失原始的结构信息。基于正则化的方法则通过直接建模示例-标记关系,显著增强了模型的表达能力与建模完整性,适合结构复杂、标记高度依赖的任务。但计算成本较高,尤其在面对大规模数据时存在效率瓶颈。表7进一步总结归纳了多示例多标记学习典型方法的优缺点。

3.2.5 流特征在线多标记学习

近些年随着社交媒体以及在线视频网站的普及,每天都会产生大量的图片和文字信息。在此类场景下随着时间的不断更新,多标记数据的特征会以流的形式逐个到达,特征空间会持续变化,且每个新流入的特征需要被立即在线处理^[145]。特征空间的动态性演化对传统的静态特征空间下的学习算法带来了诸多的挑战和问题^[146]。

特征流的在线多标记学习是多标记学习中的一种方法,它关注处理数据流中的特征,适用于在线学习的场景。如程玉胜等^[147]提出了一种动态滑动窗口加权互信息流特征选择算法 SFD-FMI-DSW,该方法通过引入补性质的模糊信息熵的方法并考虑多标记数据中的标记数据和未标记数据的互相影响,使用加权的模糊互信息度量方法结合滑动窗口的机制实现在线多标记流特征选择。白盛兴等^[148]提出了一种基于邻域粗糙集的大规模层次分类在线流特征选择算法 OHFS,克服了现有的方法忽略类别间关系和层次分类数据中类别之间的层次结构。Liu 等^[149]提出了一种基于邻域粗糙集的多标记在线流特征选择算法 OM-NRS,通过引入最大近邻准则进行示例粒化,从而将近邻粗糙集从单标记学习推广到多标记学习场景,并实现在线的多标记特征选择。后来,Liu 等^[150]针对现有的特征学习算法不能同时考虑特征的内在组结构和流特征等问题,提出了一种在线多标记组特征学习算法 OMGFS。该算法对每次流入的特征组进行逐个标记重要性评价,然后再通过考虑特征交互和特征冗余进行组间特征选择,直到所有特征流入。

虽然上述方法在不同角度上提升了流特征在线多标记学习的适应能力,但仍存在诸如计算开销高,参数选择敏感,特征流分布变化适应性不强等挑战。

表7 多示例多标记学习典型方法总结

Table 7 Summary of representative methods for multi-instance multi-label learning

算法类型	代表算法	主要机制	优势	局限性	适用场景
基于退化方法	MIMLBoost ^[140]	将问题退化为多标记学习	构建简洁,便于复用传统分	忽略标记间复杂关联;部分	样本量极小,标记关联简单和需明确决策边界的场景
	MIML-kNN ^[141]	或多实例学习任务,借助传	类模型;适用于样本量较小	方法如近邻或 Boost 受超参	
	MIMLSVM-mi ^[142]	统监督学习方法进行训练	或已知结构的任务	数影响大,训练效率较低	
基于正则化方法	MLMIL ^[143]	在模型中引入正则项建模	可同时建模标记相关性与实	训练开销大,难以在大规模	标记关联复杂,需注入领域先验,数据规模中等及鲁棒性要求高
	Multi_es ^[144]	以联合优化目标函数	例语义,适应复杂结构数据	数据中高效执行	

4 未来研究方向的思考

4.1 基于数据的内性质质的方法研究

多标记学习处理的学习问题是针对多个标记进行建模并进行预测,但是由于多标记数据的标记空间很大,而不同标记拥有的标记数据有限,从而导致了类内不平衡问题的出现。同时,每个样本拥有的平均标记数量存在差异,从而导致类间不平衡问题存在^[151]。若采用传统的分类原则处理多标记不平衡数据,往往会导致分类精度下降。目前多标记类不平衡问题的解决方案已经存在一些直接或间接的策略。一种方法是将多标记数据转化为多个二分类问题,然后通过传统的二类不平衡学习方法(如过采样、欠采样、代价敏感学习和集成学习等^[152])进行解决。另外,还存在其他策略来解决多标记类不平衡问题,如通过间接优化某些类不平衡的评价指标来实现类不平衡学习的目的^[153],通过考虑多标记数据的特性将多标记类不平衡问题转化为三类不平衡问题进行解决^[154]。此外,由于某些标记上的样本数量有限,导致模型很难学习到稳定的特征表示,并且如何在这种学习场景下去捕捉标记之间的关联同时处理类不平衡问题是极其困难的。因此需要探索新的学习策略,增强稀缺标记的特征表示,提高预测性能。

此外,在多标记场景下,不同标记的相对重要程度往往是不同的。如在自然场景图像标注中,虽然存在“海”“沙滩”和“天空”等标记,但是这些标记具体描述图像的程度是存在差异的。在电影情感分析任务中,一部影片可能包含多种基本情感元素(如喜悦、悲伤、紧张、宁静、愤怒和恐惧),通过分析这些情感元素在影片中的表现强度和相互关系,才能准确捕捉电影所传达的复杂情感体验。因此,可以看出对于描述同一个示例的多个标记之间的重要程度是不一样的。这种对标记重要程度的描述度,被称为标记分布^[155]。标记分布学习的目标对于每个样本,预测其标记的概率分布而不是直接预测单一标记。这样的方法更适应于复杂的任务,其中一个样本可能属于多个类别或存在不确定性,提供了更为丰富的信息^[156]。同时,标记分布学习在模糊性任务和细粒度分类中具有天然的优势,使其超越了传统多标记学习的技术挑战。但是,这种概率分布在真实场景中很难获得,严重依赖标注者的主观判断。因此,未来可以探索在预测标记分布的同时,衡量模型对标记关联的建模效果。

4.2 面向标记复杂性计算的方法研究

多标记数据由于其待建模类别不断扩张,标记体量呈现爆发式增长态势,导致其计算复杂性增加。相关研究已表明,利用标记关系可以有效提高多标记学习的泛化性能,并且具有很强的适用性。然而,在一些复杂背景下,学习目标可能经历变化,或者标记关系不再具备可靠性,例如标记信息源自不完整的样本标注。在这种

情境下,如何有效地利用标记关系进行学习成为未来值得深入研究的方向。此外,在以往的多标记学习研究中,一个基本的假设是:所有与示例相关的类别标记都被提供并用于训练。然而,多标记学习任务涉及的标记数量众多,而且语义相对复杂,因此对于这类任务来说,获取到完整的监督信息是非常困难的。在文本情感分析的场景中,标注者可能只标注文本中的一些关键词或短语,这些关键词或短语被认为能够反映文本的情感或主题信息。然而,这种标注方式可能导致模型忽略文本中一些模糊或罕见的表达,因为它们未被直接标注。由于标记的不完整性,模型在对待预测文本时可能会产生误导,严重影响其性能。另外,在医学图像分析领域,获取精确的图像标记通常需要大量的专业医生的时间和劳动成本,尤其是当图像涉及多个复杂的医学标记时。为处理标注数据稀缺这一问题,使用半监督多标记学习方法同时利用标注和未标注的医学图像数据进行学习。然而,这种方法在应对存在标记缺失的医学图像数据时存在一定限制,因为它假定标注的数据没有标记缺失的情况。在处理医学图像数据时,标记缺失是一种常见情况,因此半监督多标记学习方法在面对标记缺失时可能无法提供可靠的性能保证。因此,需要进一步研究和开发更为鲁棒的方法,以应对存在的标记缺失问题,提高多标记学习的实用性和准确性。

4.3 面向特征复杂性计算的方法研究

在多标记学习中,面向特征复杂性计算的方法涉及处理高维度、复杂结构的特征。维数灾难问题指的是由于标记空间的维度非常大,导致样本空间的维度爆炸,从而增加了模型训练的难度和计算的复杂性。这种情况可能引起许多传统机器学习方法的性能下降,因为它们在高维空间中需要更多的数据来进行有效的学习。特征选择可以用来降维,可以通过评估特征与标记之间的相关性,选择与标记相关性较高的特征。例如,当一个电影被标注为“科幻”和“外星人”,那么该电影被标注为“太空探险”这一类别的可能性可能会增加。相反,如果一本书被标注为“历史”这一主题,那么该书被标注为“未来科技”这一类别的可能性可能较低。同时,较为重要的是标记关系的局部性质,标记之间的关系在局部范围内的特定子集中得到体现或表达。这种局部性质可以对标记之间的依赖关系进行更精细和局部化的建模,使得模型更好地理解标记之间的关联性。

在多标记学习中,每个标记都有其自身独有的特征,这些特征对于相应标记的预测能力很强。针对每个标记强辨识力的特征进行学习,就是类属属性学习。类属属性具有区分不同标记的能力,采用联合学习标记的类属属性和标记相关性可以显著提高模型的预测性能。例如,当模型预测出某个样本属于“哺乳动物”标记时,结合“哺乳动物”与“胎生”标记之间的关联属性,就

能够更准确地判断该样本是否也属于“胎生”标记,从而显著提升多标记学习模型在复杂任务中的预测效果和泛化能力。类属性通常可以很好地反映特征空间的结构信息,例如它们的层次结构、相似性、包含关系等。因此,通过构建类属性可以实现根据标记自身信息来生成其特有类属特征,利用不同的特征空间为不同的类别标记构造分类模型,从而提高多标记学习系统的整体性能。

5 潜在的应用方向

多标记学习能很好地应对数据的复杂性、语义的关联性、标记的相关性等,在许多现实场景中已取得巨大成功,如文本分析、多媒体内容理解、生物信息学等,如图4所示。具体来说,在文本分析领域,主要任务是为文本分配一组预定义类别。由于一个文本通常同时属于多个类别,因此文本分析可以被视为多标记学习问题。常见的应用场景有工业质量检测、教育可行化学习、文献索引、标记推荐、医学诊断、专利分类、垃圾邮件过滤、文本情感分类等^[157];在多媒体内容理解领域,随着数字多媒体资源(如图像、视频、声音)的指数级增长,多媒体内容包含丰富的语义信息。得益于多标记学习在挖掘和处理复杂信息方面的卓越性,图像和视频自动标注、图像分割、人脸验证、目标识别、语音或音乐的情感检测、音乐的元数据提取等研究得到了长足的发展^[158];在生物信息学领域,通常涉及生物分子的功能、结构预测等研究。每个生物分子普遍与不止一个而是一组功能、结构相关,因此,多标记学习能广泛用于基因功能的预测和注释、蛋白质功能的预测、蛋白质亚细胞的定位和蛋白质三维结构预测等应用^[159]。另外,在其他领域,如社交网络、推荐系统、信息检索、Web挖掘、机器人导航等研究领域,多标记学习的应用也取得了不错的成果^[160]。未来,多标记学习还可以在其他潜在的应用方向发展,包括中医状态辨识、神经发育障碍和药物研发等。

中医状态辨识是在中医学理论指导下,根据生命过程中某一阶段的表征参数,辨别病位和病性等状态要素,进而概括为完整证名的思维认识过程^[161]。其中,病位是人体状态变化所反映的部位,如气血、脏腑、经络和四肢百骸等;病性是生命所处状态发生的内外平衡,阴阳偏颇和邪正斗争表现出的体征,如阴虚、阳虚、气滞、血瘀等。考虑到辨识的内容涵盖体质、生理病理特点、证等各种因素,中医状态辨识被视为一个典型的多标记学习问题^[162]。中医状态辨识这一问题的解决在机器学习、数据挖掘和人工智能等领域已有涉及,比如基于舌诊的糖尿病与糖调节受损的检测和冠心病的智能诊断及其问诊模型的研究,这些工作为从中医数据中获取潜在的应用价值和可理解模式提供了参考。但是挖掘典型疾病的局部状态模式并不是一种足够有效的方案,这是源于中医强调从整体上把握人体健康。这就要求中医状态辨识不仅需要注意利用四诊的临床症状和体征进行“四诊合参”,还要注意人体的气血、脏腑、经络和形体官窍之间的关联性,因而很有必要从人体自身、人与社会和人与自然等各种“整体”角度来进行状态辨识。

神经发育障碍是在发育阶段起病的一类疾病,影响大脑的生长和发育,一般出现在发育早期,主要表现为人际交往障碍、交流沟通障碍以及兴趣和行为方面的异常。神经发育障碍为分类上的大类谱系,具体障碍包括孤独症谱系障碍、智力障碍、交流障碍、注意力缺陷多动障碍和运动障碍等。其中,孤独症谱系障碍是神经发育障碍中具有代表性的疾病之一,又包含孤独症、阿斯伯格综合征、童年瓦解症以及未特定的广泛性发育障碍等多个分类^[9]。可以看出,神经发育障碍同样是一个多标记学习的典型应用,而且标记的组织具有层次结构的形式。目前,这类疾病的成因尚不明确,从病理学角度来说,与遗传、环境、大脑过度发育以及神经炎等因素相关。已有一些工作尝试利用大脑的核磁共振图像来进行疾病预测,也有一些工作通过分析脑发育过程中

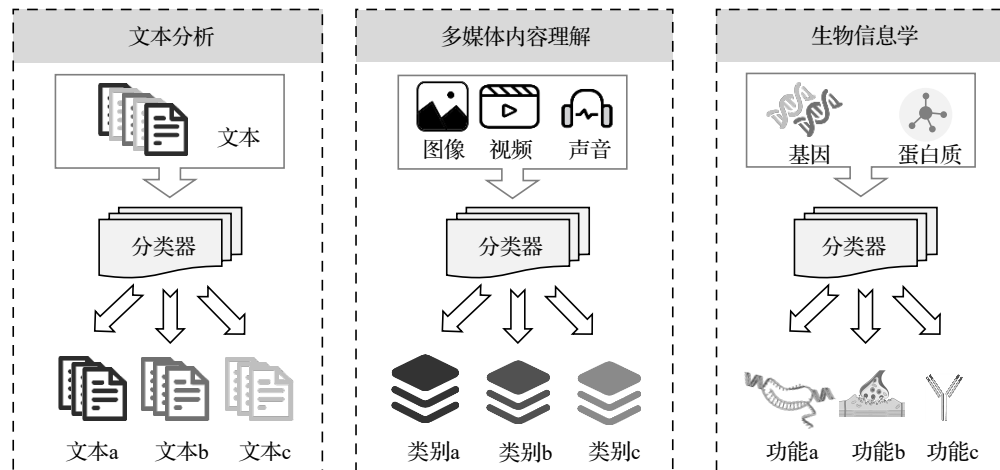


图4 多标记学习应用场景

Fig.4 Multi-label learning application scenarios

lncRNA 基因表达谱来开展相关研究。

全新结构的药物研发周期长、耗资大、风险高,加快创新药物的问世已经成为当前亟待解决的医学难题^[63]。多标记学习技术的融入,正为创新药物研发注入新的活力。在创新研发的早期阶段,药物-靶点相互作用预测是核心环节之一。传统方法往往针对单一靶点筛选化合物,忽略了候选药物可能与多个潜在靶点存在的关联。许多疾病(如癌症、自身免疫性疾病)具有复杂的病理机制,涉及多靶点协同作用,单一靶点药物往往难以达到理想疗效。多标记学习技术能够整合药物分子结构、靶点蛋白质序列、基因表达谱等多源数据,同时预测一种药物与多个靶点的相互作用关系^[64]。例如,在抗肿瘤药物研发中,它可以通过学习已知抗肿瘤药物与激酶靶点家族的作用模式,一次性预测候选化合物对多个激酶靶点的抑制活性,帮助研发人员快速锁定具有多靶点协同效应的先导化合物,大幅减少体外实验的筛选范围,降低早期研发成本。此外,创新研发不仅包括全新化合物的开发,还涵盖对已上市药物或临床阶段药物的新适应症挖掘^[65]。最著名的案例便是西地那非,该药最初用于治疗高血压、心绞痛等心血管疾病,后来发现它可以治疗男性勃起功能障碍。而多标记学习技术能进一步挖掘药物与疾病的隐性关联,将药物的化学结构、药理特征、副作用谱等作为输入特征,将多种疾病作为标记,通过构建多标记分类模型,识别药物与不同疾病之间未被发现的联系。对于处于临床II期或III期失败的候选药物,多标记学习能重新分析其与其他疾病的潜在匹配度,这在创新研发资源有限的背景下,能显著提高研发成功率。例如,某款在阿尔茨海默病临床实验中失败的药物,通过多标记学习模型分析其对神经炎症、突触修复等相关疾病的作用潜力,可能被重新定位为帕金森病辅助治疗药物,从而挽救前期巨额研发投入^[66]。

此外,多标记学习在脑电信号情感识别^[67]、人机协作^[68]、自动驾驶路径规划^[69]等领域也表现出令人期待的发展趋势。在脑电信号情感识别领域,情感识别通常涉及多个情感维度,每个情感维度可以被视为一个标记。目前有关文本、图像、音频的多标记情感分析已受到广泛关注,然而通过脑电信号进行多标记情感识别研究仍具有挑战性;在人机协作领域,旨在开发为在执行体力任务的人类劳动者提供帮助的机器人以提升工作效率。为此,准确且全面的任务辅助预测成为机器人研发的关键。例如,临床机器人助手通过预测外科医生的意图来传递合适的手术器械协助手术过程。考虑到进行特定手术操作时需要同时使用多种器械,因此任务辅助预测可以被视为多标记学习问题;在自动驾驶路径规划领域,以驾驶体验为例,可以将路径规划问题分解为安全性、效率性和舒适性等多种驾驶体验,每种体验对应一个标记。同时考虑多种体验进行路径规划旨在通

过权衡不同驾驶者的多方面关切点,提供一个较为全面的规划方案。

综上,多标记学习已成为解决现实世界复杂问题和复杂数据的有效工具。其灵活性和优越的信息处理能力使其在许多现实场景中得到了广泛的应用,并为未来的研究提供了广阔的前景。

6 结束语

多标记学习是一个非常具有前景性的研究领域,引起了国内外学者的广泛关注。目前,已经产生了许多标志性的成果,这些成果不仅展示了多标记学习的潜力和应用价值,也为未来的研究提供了重要的指导和借鉴。随着技术的不断发展和数据的不断增长,多标记学习的研究将会更加深入和广泛,有望在更多的领域得到应用和推广。但是,目前关于复杂数据背景下的多标记学习,仍存在很多关键科学问题亟需解决。为了总结目前研究的趋势和热点研究,探讨该方向的未来研究内容,本文系统地就目前研究的主要学习方法和应用成果进行了归纳、总结和分析,同时指出了当前研究面临的挑战和未来的研究方向。随着研究的深入进行,复杂数据背景下的多标记学习算法将会更加成熟,并且在更多潜在应用领域发挥作用。

参考文献:

- [1] 桑江徽,姜海燕.基于联合分布的多标记迁移学习[J].计算机工程与应用,2021,57(9):154-161.
SANG J H, JIANG H Y. Multi-label transfer learning algorithm based on joint distribution alignment[J]. Computer Engineering and Applications, 2021, 57(9): 154-161.
- [2] YAN Y F, HUANG S J, CHEN S Y, et al. Active learning with query generation for cost-effective text classification[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 6583-6590.
- [3] TANG C, LIU X W, WANG P C, et al. Adaptive hypergraph embedded semi-supervised multi-label image annotation[J]. IEEE Transactions on Multimedia, 2019, 21(11): 2837-2849.
- [4] HUANG Z A, ZHANG J, ZHU Z X, et al. Identification of autistic risk candidate genes and toxic chemicals via multilabel learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(9): 3971-3984.
- [5] DAI J H, CHEN W X, QIAN Y H. Multi-label feature selection with missing features via implicit label replenishment and positive correlation feature recovery[J]. IEEE Transactions on Knowledge and Data Engineering, 2025, 37(4): 2042-2055.
- [6] LI Y T, ZHANG J, WU H R, et al. Consistent and specific multi-view multi-label learning with correlation information [J]. Information Sciences, 2025, 687: 121395.
- [7] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-

- label data[M]//MAIMON O, ROKACH L.Data mining and knowledge discovery handbook.Boston, MA: Springer, 2009: 667-685.
- [8] ZHANG M L, ZHOU Z H.A review on multi-label learning algorithms[J].IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [9] GIBAJA E, VENTURA S.A tutorial on multilabel learning[J].ACM Computing Surveys, 2015, 47(3): 1-38.
- [10] 余鹰.多标记学习研究综述[J].计算机工程与应用, 2015, 51(17): 20-27.
- YU Y.Survey on multi-label learning[J].Computer Engineering and Applications, 2015, 51(17): 20-27.
- [11] LIU W W, WANG H B, SHEN X B, et al.The emerging trends of multi-label learning[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7955-7974.
- [12] WU X Z, ZHOU Z H.A unified view of multi-label performance measures[C]//Proceedings of the 34th International Conference on Machine Learning, 2017: 3780-3788.
- [13] SCHAPIRE R E, SINGER Y.BoosTexter: a boosting-based system for text categorization[J].Machine Learning, 2000, 39(2): 135-168.
- [14] DEMBCZYŃSKI K, WAEGEMAN W, CHENG W W, et al.Regret analysis for performance metrics in multi-label classification: the case of hamming and subset zero-one loss [C]//Proceedings of the 21st European Conference on Machine Learning and Knowledge Discovery in Databases.Berlin: Springer, 2010: 280-295.
- [15] ZHOU Z Y, WANG A, XIA S J, et al.Towards traditional Chinese medicine symptom prediction with multi-label learning method[C]//Proceedings of the 2024 14th International Conference on Information Technology in Medicine and Education.Piscataway: IEEE, 2025: 1110-1114.
- [16] ALALGA A, BENABDESLEM K, MANSOURI D E K.3-3FS: ensemble method for semi-supervised multi-label feature selection[J].Knowledge and Information Systems, 2021, 63(11): 2969-2999.
- [17] DAI L, DU G D, ZHANG J, et al.Joint multilabel classification and feature selection based on deep canonical correlation analysis[J].Concurrency and Computation: Practice and Experience, 2020, 32(22): e5864.
- [18] LIU W W, TSANG I W.On the optimality of classifier chain for multi-label classification[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems, 2015: 712-720.
- [19] ZHANG M L, WU L.Lift: multi-label learning with label-specific features[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107-120.
- [20] ZHAO T N, ZHANG Y J, MIAO D Q, et al.Fine-grained local label correlation for multi-label classification[J].Knowledge-Based Systems, 2025, 314: 113210.
- [21] ZHOU A, LIU B, WANG J, et al.Batch selection for multi-label classification guided by uncertainty and dynamic label correlations[C]//Proceedings of the 39th AAAI Conference on Artificial Intelligence.Palo Alto: AAAI, 2025: 22902-22909.
- [22] JIA Q W, DENG T Q, YANG M, et al.Class label fusion guided correlation learning for incomplete multi-label classification[J].Information Fusion, 2025, 120: 103072.
- [23] BOUTELL M R, LUO J B, SHEN X P, et al.Learning multi-label scene classification[J].Pattern Recognition, 2004, 37(9): 1757-1771.
- [24] CLARE A, KING R D.Knowledge discovery in multi-label phenotype data[C]//Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery.Berlin: Springer, 2001: 42-53.
- [25] TSOUMAKAS G, VLAHAVAS I.Random k-labelsets: an ensemble method for multilabel classification[C]//Proceedings of the 18th European Conference on Machine Learning.Berlin: Springer, 2007: 406-417.
- [26] READ J, PFAHRINGER B, HOLMES G, et al.Classifier chains for multi-label classification[J].Machine Learning, 2011, 85(3): 333-359.
- [27] HUANG J, LI G R, HUANG Q M, et al.Learning label-specific features and class-dependent labels for multi-label classification[J].IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309-3323.
- [28] LIN Y J, HU Q H, LIU J H, et al.Streaming feature selection for multilabel learning based on fuzzy mutual information[J].IEEE Transactions on Fuzzy Systems, 2017, 25(6): 1491-1507.
- [29] JIA Q W, DENG T Q, ZHANG Z A, et al.Pointwise fuzzy mutual information based multi-label feature selection via feature low-rank regularization[J].Applied Soft Computing, 2025, 179: 113301.
- [30] ZHU Y, KWOK J T, ZHOU Z H.Multi-label learning with global and local label correlation[J].IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1081-1094.
- [31] HOU P, GENG X, ZHANG M L.Multi-label manifold learning[C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence.Palo Alto: AAAI, 2016: 1680-1686.
- [32] LIU B, LI Y M, XU Z L.Manifold regularized matrix completion for multi-label learning with ADMM[J].Neural Networks, 2018, 101: 57-67.
- [33] NGUYEN V A, BOYD-GRABER J, RESNIK P, et al.Learning a concept hierarchy from multi-labeled documents [C]//Advances in Neural Information Processing Systems 27, 2014: 3671-3679.
- [34] GUO Y H, SCHUURMANS D.Multi-label classification with output kernels[C]//Proceedings of the 24th European Conference on Machine Learning and Knowledge Discovery in Databases.Berlin: Springer, 2013: 417-432.
- [35] AKBARNEJAD A H, BAGHSHAH M S.An efficient semi-

- supervised multi-label classifier capable of handling missing labels[J].IEEE Transactions on Knowledge and Data Engineering, 2019, 31(2): 229-242.
- [36] ELISSEEFF A, WESTON J.A kernel method for multi-labelled classification[C]//Advances in Neural Information Processing Systems 14, 2001: 681-687.
- [37] XIE S H, KONG X N, GAO J, et al.Multilabel consensus classification[C]//Proceedings of the 2013 IEEE 13th International Conference on Data Mining.Piscataway: IEEE, 2014: 1241-1246.
- [38] CORTES C, MOHRI M.AUC optimization vs.error rate minimization[C]//Advances in Neural Information Processing Systems 16, 2003: 313-320.
- [39] WEI T, GUO L Z, LI Y F, et al.Learning safe multi-label prediction for weakly labeled data[J].Machine Learning, 2018, 107(4): 703-725.
- [40] BI W, KWOK J T.Bayes-optimal hierarchical multilabel classification[J].IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11): 2907-2918.
- [41] WU B Y, JIA F, LIU W, et al.Multi-label learning with missing labels using mixed dependency graphs[J].International Journal of Computer Vision, 2018, 126(8): 875-896.
- [42] WEHRMANN J, CERRI R, BARROS R.Hierarchical multi-label classification networks[C]//Proceedings of the 35th International Conference on Machine Learning, 2018: 5225-5234.
- [43] HUANG S J, ZHOU Z H.Multi-label learning by exploiting label correlations locally[C]//Proceedings of the 26th AAAI Conference on Artificial Intelligence.Palo Alto: AAAI, 2012: 949-955.
- [44] ROWEIS S T, SAUL L K.Nonlinear dimensionality reduction by locally linear embedding[J].Science, 2000, 290: 2323-2326.
- [45] XU M, JIN R, ZHOU Z H.Speedup matrix completion with side information: application to multi-label learning[C]//Advances in Neural Information Processing Systems 26, 2013: 2301-2309.
- [46] CHE X Y, CHEN D G, DENG J, et al.Exploiting local label correlation from sample perspective for multi-label classification via three-way decision theory[J].Applied Soft Computing, 2023, 149: 110950.
- [47] WERTZ L, BOGOJESKA J, MIRYLENKA K, et al.Reinforced active learning for low-resource, domain-specific, multi-label text classification[C]//Findings of the Association for Computational Linguistics: ACL 2023.Stroudsburg: ACL, 2023: 10959-10977.
- [48] WEI T, LI Y F.Learning compact model for large-scale multi-label data[C]//Proceedings of the 33th AAAI Conference on Artificial Intelligence.Palo Alto: AAAI, 2019: 5385-5392.
- [49] ZHANG Y, WANG Y, LIU X Y, et al.Large-scale multi-label classification using unknown streaming images[J].Pattern Recognition, 2020, 99: 107100.
- [50] LIU C L, WEN J, XU Y, et al.Reliable representation learning for incomplete multi-view missing multi-label classification [J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(6): 4940-4956.
- [51] DU G D, ZHANG J, ZHANG N, et al.Semi-supervised imbalanced multi-label classification with label propagation[J].Pattern Recognition, 2024, 150: 110358.
- [52] NGUYEN T T, NGUYEN T T T, LUONG A V, et al.Multi-label classification via label correlation and first order feature dependance in a data stream[J].Pattern Recognition, 2019, 90: 35-51.
- [53] LIU J H, LIN Y J, DING W P, et al.Fuzzy mutual information-based multilabel feature selection with label dependency and streaming labels[J].IEEE Transactions on Fuzzy Systems, 2023, 31(1): 77-91.
- [54] DU J, VONG C M.Robust online multilabel learning under dynamic changes in data distribution with labels[J].IEEE Transactions on Cybernetics, 2020, 50(1): 374-385.
- [55] ZHANG N, WANG A, LU P W, et al.Multi-label feature selection with feature reconstruction and label correlations [J].Expert Systems with Applications, 2025, 285: 127993.
- [56] YANG Y, FU Z Y, ZHAN D C, et al.Semi-supervised multimodal multi-instance multi-label deep network with optimal transport[J].IEEE Transactions on Knowledge and Data Engineering, 2021, 33(2): 696-709.
- [57] HOU B J, ZHANG L J, ZHOU Z H.Learning with feature evolvable streams[J].IEEE Transactions on Knowledge and Data Engineering, 2021, 33(6): 2602-2615.
- [58] GE Z Q, GUAN Y Y, LI X M, et al.Consistent, balanced, and overlapping label trees for extreme multi-label learning[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management.New York: ACM, 2022: 551-560.
- [59] 刘慧婷,冷新杨,王利利,等.联合嵌入式多标签分类算法[J].自动化学报,2019,45(10): 1969-1982.
- LIU H T, LENG X Y, WANG L L, et al.A joint embedded multi-label classification algorithm[J].Acta Automatica Sinica, 2019, 45(10): 1969-1982.
- [60] LIANG D J, ZHANG H X, YUAN D F, et al.Multi-head encoding for extreme label classification[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(3): 2199-2211.
- [61] JASINSKA K, DEMBCZYNSKI K, BUSA-FEKETE R, et al.Extreme F-measure maximization using sparse probability estimates[C]//Proceedings of the 33rd International Conference on Machine Learning, 2016: 1435-1444.
- [62] ZHANG J, CHANG W C, YU H F, et al.Fast multi-resolution transformer fine-tuning for extreme multi-label text classification[C]//Advances in Neural Information Processing Systems 34, 2021: 7267-7280.

- [63] JIANG T, WANG D Q, SUN L L, et al. LightXML: Transformer with dynamic negative sampling for high-performance extreme multi-label text classification[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021: 7987-7994.
- [64] CAI Y R, XIE J R, TANG B S, et al. First-order multi-label learning with cross-modal interactions for multimodal emotion recognition[C]//Proceedings of the 1st International Workshop on Multimodal and Responsible Affective Computing. New York: ACM, 2023: 13-20.
- [65] BHATIA K, JAIN H, KAR P, et al. Sparse local embeddings for extreme multi-label classification[C]//Advances in Neural Information Processing Systems 28, 2015: 730-738.
- [66] XU C, TAO D C, XU C. Robust extreme multi-label learning[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1275-1284.
- [67] PRABHU Y, VARMA M. FastXML: a fast, accurate and stable tree-classifier for extreme multi-label learning[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 263-272.
- [68] SI S, ZHANG H, KEERTHI S S, et al. Gradient boosted decision trees for high dimensional sparse output[C]//Proceedings of the 34th International Conference on Machine Learning, 2017: 3182-3190.
- [69] JAIN H, PRABHU Y, VARMA M. Extreme multi-label loss functions for recommendation, tagging, ranking & other missing label applications[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 935-944.
- [70] YEN I E H, HUANG X R, DAI W, et al. PPDsparse: a parallel primal-dual sparse method for extreme classification[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 545-553.
- [71] WEI T, LI Y F. Does tail label help for large-scale multi-label learning?[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2315-2324.
- [72] BI W, KWOK J T. Multi-label classification on tree and DAG-structured hierarchies[C]//Proceedings of the 28th International Conference on International Conference on Machine Learning, 2011: 17-24.
- [73] BALASUBRAMANIAN K, LEBANON G. The landmark selection method for multiple output prediction[C]//Proceedings of the 29th International Conference on International Conference on Machine Learning, 2012: 283-290.
- [74] TAN A H, LIANG J Y, WU W Z, et al. Semi-supervised partial multi-label classification via consistency learning[J]. Pattern Recognition, 2022, 131: 108839.
- [75] JING L P, YANG L, YU J, et al. Semi-supervised low-rank mapping learning for multi-label classification[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1483-1491.
- [76] DONG H C, LI Y F, ZHOU Z H. Learning from semi-supervised weak-label data[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence and 13th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI, 2018: 2926-2933.
- [77] WU L, ZHANG M L. Multi-label classification with unlabeled data: an inductive approach[C]//Proceedings of the 5th Asian Conference on Machine Learning, 2013: 197-212.
- [78] KONG X N, NG M K, ZHOU Z H. Transductive multilabel learning via label set propagation[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(3): 704-719.
- [79] XIONG L M, ZHOU J, ZHU Q X, et al. A confidence-based partial label learning model for crowd-annotated named entity recognition[C]//Findings of the Association for Computational Linguistics: ACL 2023. Stroudsburg: ACL, 2023: 1375-1386.
- [80] WANG H B, YANG S S, LYU G Y, et al. Deep partial multi-label learning with graph disambiguation[C]//Proceedings of the 32nd International Joint Conference on Artificial Intelligence, 2023: 4308-4316.
- [81] XU L L, WANG Z, SHEN Z F, et al. Learning low-rank label correlations for multi-label classification with missing labels[C]//Proceedings of the 2014 IEEE International Conference on Data Mining. Piscataway: IEEE, 2015: 1067-1072.
- [82] WU B Y, LYU S W, GHANEM B. Constrained submodular minimization for missing labels and class imbalance in multi-label learning[C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2016: 2229-2236.
- [83] ZHAN W, ZHANG M L. Inductive semi-supervised multi-label learning with co-training[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 1305-1314.
- [84] WANG L C, LIU Y Y, QIN C, et al. Dual relation semi-supervised multi-label learning[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 6227-6234.
- [85] FENG L, AN B, HE S. Collaboration based multi-label learning[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence and 31st Innovative Applications of Artificial Intelligence Conference and 9th AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI, 2019: 3550-3557.
- [86] SONG Z X, MENG Z Q, ZHANG Y F, et al. Semi-supervised multi-label learning for graph-structured data[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM, 2021: 1723-1733.

- [87] JI Z, CUI B Y, LI H H, et al. Deep ranking for image zero-shot multi-label classification[J]. *IEEE Transactions on Image Processing*, 2020, 29: 6549-6560.
- [88] PALATUCCI M, POMERLEAU D, HINTON G, et al. Zero-shot learning with semantic output codes[C]// *Advances in Neural Information Processing Systems 22*, 2009: 1410-1418.
- [89] KADAM S, VAIDYA V. Review and analysis of zero, one and few shot learning approaches[C]// *Proceedings of the 18th International Conference on Intelligent Systems Design and Applications*. Cham: Springer, 2020: 100-112.
- [90] FU Y W, YANG Y X, HOSPEDALES T, et al. Transductive multi-label zero-shot learning[C]// *Proceedings of the British Machine Vision Conference 2014*, 2014: 1-11.
- [91] MENSINK T, GAVVES E, SNOEK C G M. COSTA: co-occurrence statistics for zero-shot classification[C]// *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2014: 2441-2448.
- [92] GONG Y C, KE Q F, ISARD M, et al. A multi-view embedding space for modeling Internet images, tags, and their semantics[J]. *International Journal of Computer Vision*, 2014, 106(2): 210-233.
- [93] WANG Q, CHEN K. Zero-shot visual recognition via bidirectional latent embedding[J]. *International Journal of Computer Vision*, 2017, 124(3): 356-383.
- [94] ZHANG Y, GONG B Q, SHAH M. Fast zero-shot image tagging[C]// *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2016: 5985-5994.
- [95] REN Z, JIN H L, LIN Z, et al. Multi-instance visual-semantic embedding[C]// *Proceedings of the British Machine Vision Conference*, 2017.
- [96] CHANGPINYO S, CHAO W L, SHA F. Predicting visual exemplars of unseen classes for zero-shot learning[C]// *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2017: 3496-3505.
- [97] NOROUZI M, MIKOLOV T, BENGIO S, et al. Zero-shot learning by convex combination of semantic embeddings [C]// *Proceedings of the 2nd International Conference on Learning Representations*, 2014.
- [98] ZOU L G, ZHOU T, DAI J H. Online multilabel streaming feature selection by label enhancement and fuzzy synergistic discrimination information[J]. *IEEE Transactions on Fuzzy Systems*, 2025, 33(7): 2203-2214.
- [99] LIN Y J, HU Q H, ZHANG J, et al. Multi-label feature selection with streaming labels[J]. *Information Sciences*, 2016, 372: 256-275.
- [100] PHAM A, RAICH R, FERN X, et al. Multi-instance multi-label learning in the presence of novel class instances[C]// *Proceedings of the 32nd International Conference on Machine Learning*, 2015: 2427-2435.
- [101] ZHU Y, TING K M, ZHOU Z H. Discover multiple novel labels in multi-instance multi-label learning[C]// *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI, 2017: 2977-2984.
- [102] ZHU Y, TING K M, ZHOU Z H. Multi-label learning with emerging new labels[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(10): 1901-1914.
- [103] LIU J H, LI Y W, WENG W, et al. Feature selection for multi-label learning with streaming label[J]. *Neurocomputing*, 2020, 387: 268-278.
- [104] WANG Y, WANG C Z, DENG T Q, et al. Multi-label feature selection via nonlinear mapping and manifold regularization[J]. *Information Sciences*, 2025, 704: 121965.
- [105] TIBSHIRANI R. Regression shrinkage and selection via the lasso: a retrospective[J]. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 2011, 73(3): 273-282.
- [106] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $\ell_2, 1$ -norms minimization[C]// *Advances in Neural Information Processing Systems 23*, 2010: 1813-1821.
- [107] WANG J, WEI J M, YANG Z L. Supervised feature selection by preserving class correlation[C]// *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. New York: ACM, 2016: 1613-1622.
- [108] LEE J, KIM D W. Feature selection for multi-label classification using multivariate mutual information[J]. *Pattern Recognition Letters*, 2013, 34(3): 349-357.
- [109] LEE J, KIM D W. Fast multi-label feature selection based on information-theoretic feature ranking[J]. *Pattern Recognition*, 2015, 48(9): 2761-2771.
- [110] LIM H, LEE J, KIM D W. Optimization approach for feature selection in multi-label classification[J]. *Pattern Recognition Letters*, 2017, 89: 25-30.
- [111] ZHANG Y, ZHOU Z H. Multilabel dimensionality reduction via dependence maximization[J]. *ACM Transactions on Knowledge Discovery from Data*, 2010, 4(3): 1-21.
- [112] ZHANG M L, PEÑA J M, ROBLES V. Feature selection for multi-label naive Bayes classification[J]. *Information Sciences*, 2009, 179(19): 3218-3229.
- [113] YAN P, LI Y. Graph-margin based multi-label feature selection [C]// *Proceedings of the 27th European Conference on Machine Learning and Knowledge Discovery in Databases*. Cham: Springer, 2016: 540-555.
- [114] HUANG R, JIANG W D, SUN G L. Manifold-based constraint Laplacian score for multi-label feature selection[J]. *Pattern Recognition Letters*, 2018, 112: 346-352.
- [115] ZHANG J, LUO Z M, LI C D, et al. Manifold regularized discriminative feature selection for multi-label learning[J]. *Pattern Recognition*, 2019, 95: 136-150.
- [116] ZHANG J, WU H R, JIANG M, et al. Group-preserving label-specific feature selection for multi-label learning[J]. *Expert*

- Systems with Applications, 2023, 213: 118861.
- [117] TAN Y, SUN D, SHI Y, et al. Bi-directional mapping for multi-label learning of label-specific features[J]. Applied Intelligence, 2022, 52(7): 8147-8166.
- [118] HUANG J, LI G R, HUANG Q M, et al. Joint feature selection and classification for multilabel learning[J]. IEEE Transactions on Cybernetics, 2018, 48(3): 876-889.
- [119] LIN Y J, HU Q H, LIU J H, et al. Multi-label feature selection based on max-dependency and Min-redundancy[J]. Neurocomputing, 2015, 168: 92-103.
- [120] JIAN L, LI J D, SHU K, et al. Multi-label informed feature selection[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 1627-1633.
- [121] LI L, LIU H W, MA Z J, et al. Multi-label feature selection via information gain[C]//Proceedings of the 10th International Conference on Advanced Data Mining and Applications. Cham: Springer 2014: 345-355.
- [122] ZHANG P, LIU G X, GAO W F. Distinguishing two types of labels for multi-label feature selection[J]. Pattern Recognition, 2019, 95: 72-82.
- [123] ZHANG P, SHENG J Y, GAO W F, et al. Multi-label feature selection method based on dynamic weight[J]. Soft Computing, 2022, 26(6): 2793-2805.
- [124] GAO W F, HAO P T, WU Y, et al. A unified low-order information-theoretic feature selection framework for multi-label learning[J]. Pattern Recognition, 2023, 134: 109111.
- [125] YANG Y, CHEN H M, MI Y, et al. Multi-label feature selection based on stable label relevance and label-specific features[J]. Information Sciences, 2023, 648: 119525.
- [126] XU S P, YANG X B, YU H L, et al. Multi-label learning with label-specific feature reduction[J]. Knowledge-Based Systems, 2016, 104: 52-61.
- [127] GUO Y M, CHUNG F, LI G Z, et al. Leveraging label-specific discriminant mapping features for multi-label learning[J]. ACM Transactions on Knowledge Discovery from Data, 2019, 13(2): 1-23.
- [128] HUANG J, LI G R, HUANG Q M, et al. Learning label specific features for multi-label classification[C]//Proceedings of the 2015 IEEE International Conference on Data Mining. Piscataway: IEEE, 2016: 181-190.
- [129] KASHEF S, NEZAMABADI-POUR H. A label-specific multi-label feature selection algorithm based on the Pareto dominance concept[J]. Pattern Recognition, 2019, 88: 654-667.
- [130] ZHANG J, LI C D, CAO D L, et al. Multi-label learning with label-specific features by resolving label correlations[J]. Knowledge-Based Systems, 2018, 159: 148-157.
- [131] HUANG M, ZHUANG F Z, ZHANG X, et al. Supervised representation learning for multi-label classification[J]. Machine Learning, 2019, 108(5): 747-763.
- [132] ZHANG M L, ZHOU Z H. Multilabel neural networks with applications to functional genomics and text categorization[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338-1351.
- [133] NAM J, KIM J, LOZA MENCÍA E, et al. Large-scale multi-label text classification: revisiting neural networks[C]//Proceedings of the 25th European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2014: 437-452.
- [134] YE H C K, WU W C, KO W J, et al. Learning deep latent space for multi-label classification[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2017: 2838-2844.
- [135] WANG K X, YANG M, YANG W Q, et al. Deep correlation structure preserved label space embedding for multi-label classification[C]//Proceedings of the 10th Asian Conference on Machine Learning, 2018: 1-16.
- [136] SHEN X B, LIU W W, LUO Y, et al. Deep binary prototype multi-label learning[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. New York: ACM, 2018: 2675-2681.
- [137] WANG J, YANG Y, MAO J H, et al. CNN-RNN: a unified framework for multi-label image classification[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 2285-2294.
- [138] CHEN Z M, WEI X S, WANG P, et al. Multi-label image recognition with graph convolutional networks[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 5172-5181.
- [139] YANG L, WU X Z, JIANG Y, et al. Multi-label learning with deep forest[C]//Proceedings of the 24th European Conference on Artificial Intelligence, 2020: 1634-1641.
- [140] ZHOU Z H, ZHANG M L. Multi-instance multi-label learning with application to scene classification[C]//Advances in Neural Information Processing Systems 19, 2007: 1609-1616.
- [141] ZHANG M L. A k -nearest neighbor based multi-instance multi-label learning algorithm[C]//Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence. Piscataway: IEEE, 2010: 207-212.
- [142] ZHOU Z H, ZHANG M L, HUANG S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1): 2291-2320.
- [143] ZHA Z J, HUA X S, MEI T, et al. Joint multi-label multi-instance learning for image classification[C]//Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.
- [144] LAUSER B, HOTH O A. Automatic multi-label subject indexing in a multilingual environment[C]//Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries. Berlin: Springer, 2003: 140-151.

- [145] WU X D, YU K, DING W, et al. Online feature selection with streaming features[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(5): 1178-1192.
- [146] YU K, WU X D, DING W, et al. Scalable and accurate online feature selection for big data[J]. *ACM Transactions on Knowledge Discovery from Data*, 2016, 11(2): 1-39.
- [147] 程玉胜, 李雨, 王一宾, 等. 动态滑动窗口加权互信息流特征选择[J]. *南京大学学报(自然科学)*, 2018, 54(5): 974-985.
- CHENG Y S, LI Y, WANG Y B, et al. Streaming feature selection with weighted fuzzy mutual information based on dynamic sliding window[J]. *Journal of Nanjing University (Natural Science)*, 2018, 54(5): 974-985.
- [148] 白盛兴, 林耀进, 王晨曦, 等. 基于邻域粗糙集的大规模层次分类在线流特征选择[J]. *模式识别与人工智能*, 2019, 32(9): 811-820.
- BAI S X, LIN Y J, WANG C X, et al. Large-scale hierarchical classification online streaming feature selection based on neighborhood rough set[J]. *Pattern Recognition and Artificial Intelligence*, 2019, 32(9): 811-820.
- [149] LIU J H, LIN Y J, LI Y W, et al. Online multi-label streaming feature selection based on neighborhood rough set[J]. *Pattern Recognition*, 2018, 84: 273-287.
- [150] LIU J H, LIN Y J, WU S X, et al. Online multi-label group feature selection[J]. *Knowledge-Based Systems*, 2018, 143: 42-57.
- [151] BRAYTEE A, LIU W, ANAISSI A, et al. Correlated multi-label classification with incomplete label space and class imbalance[J]. *ACM Transactions on Intelligent Systems and Technology*, 2019, 10(5): 1-26.
- [152] DU G D, ZHANG J, JIANG M, et al. Graph-based class-imbalance learning with label enhancement[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(9): 6081-6095.
- [153] DEMBCZYNSKI K, JACHNIK A, KOTLOWSKI W, et al. Optimizing the F-measure in multi-label classification: plug-in rule approach versus structured loss minimization [C]//*Proceedings of the 30th International Conference on Machine Learning*, 2013: 1130-1138.
- [154] ZHANG M L, LI Y K, YANG H, et al. Towards class-imbalance aware multi-label learning[J]. *IEEE Transactions on Cybernetics*, 2022, 52(6): 4459-4471.
- [155] GENG X. Label distribution learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(7): 1734-1748.
- [156] XU N, LIU Y P, GENG X. Label enhancement for label distribution learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(4): 1632-1643.
- [157] LIU Z, HUANG Y J, XIA X C, et al. All is attention for multi-label text classification[J]. *Knowledge and Information Systems*, 2025, 67(2): 1249-1270.
- [158] PACHET F, ROY P. Improving multilabel analysis of music titles: a large-scale validation of the correction approach [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2009, 17(2): 335-343.
- [159] ZHAO W Z, LUO S J, WU H F, et al. A multi-label learning framework for predicting antibiotic resistance genes via dual-view modeling[J]. *Briefings in Bioinformatics*, 2022, 23(3): 1-12.
- [160] QIAN W B, HUANG J T, XU F K, et al. A survey on multi-label feature selection from perspectives of label fusion[J]. *Information Fusion*, 2023, 100: 101948.
- [161] DAI L, ZHANG J, LI C D, et al. Multi-label feature selection with application to TCM state identification[J]. *Concurrency and Computation: Practice and Experience*, 2019, 31(23): e4634.
- [162] WENG H, LIU Z Q, MAXWELL A, et al. Multi-label symptom analysis and modeling of TCM diagnosis of hypertension[C]//*Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine*. Piscataway: IEEE, 2019: 1922-1929.
- [163] ZHANG J, LI C D, LIN Y J, et al. Computational drug repositioning using collaborative filtering via multi-source fusion [J]. *Expert Systems with Applications*, 2017, 84: 281-289.
- [164] GUAN R C, WANG X, YANG M Q, et al. Multi-label deep learning for gene function annotation in cancer pathways [J]. *Scientific Reports*, 2018, 8: 267.
- [165] ZHU X Y, LU W M. Multi-label classification with dual tail-node augmentation for drug repositioning[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2023, 20(5): 3068-3079.
- [166] ZHANG W, LIU F, LUO L Q, et al. Predicting drug side effects by multi-label learning and ensemble learning[J]. *BMC Bioinformatics*, 2015, 16(1): 365.
- [167] ZHONG P X, WANG D, MIAO C Y. EEG-based emotion recognition using regularized graph neural networks[J]. *IEEE Transactions on Affective Computing*, 2022, 13(3): 1290-1301.
- [168] PETERNEL L, FANG C, TSAGARAKIS N, et al. A selective muscle fatigue management approach to ergonomic human-robot co-manipulation[J]. *Robotics and Computer-Integrated Manufacturing*, 2019, 58: 69-79.
- [169] JIAN Z Q, CHEN S T, ZHANG S Y, et al. Multi-model-based local path planning methodology for autonomous driving: an integrated framework[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(5): 4187-4200.