

文章编号:1671-5896(2026)04-0851-12

基于 Transformer 的半监督偏多标签图像分类方法

李 艳^{1,2}, 周志龙³, 祝 彪⁴

(1. 北京师范大学 文理学院, 广东 珠海 519085; 2. 北京师范大学珠海分校 应用数学学院, 广东 珠海 519087;
3. 北京师范大学 统计学院, 北京 100875; 4. 河北大学 数学与信息科学学院, 河北 保定 071002)

摘要: 针对在半监督偏多标签学习 (SSPML: Semi-Supervised Partial Multi-Label Learning) 方法中, 特征和标签空间之间以及各自内部存在复杂的相关性, 且利用标签相关性构建模型时, 容易受到候选标签集中噪声标签干扰的问题, 基于 Transformer 提出一种新的半监督偏多标签图像分类算法 T-SSPML (Transformer Encoder-Based Semi-Supervised Partial Multi-Label Learning)。利用 Transformer 编码器构建主干网络, 将图像和标签特征共同作为输入, 并定义新的损失函数, 利用多头自注意力机制学习整体性结构信息以及标签间的关联信息。同时采用标签掩码机制, 提升模型学习标签关联信息的能力。通过在多个数据集上实验结果表明, T-SSPML 算法相比其他对比算法在绝大多数情况下都取得了更好的分类性能。

关键词: 偏多标签学习; 半监督学习; 图像分类; 一致性正则化; 伪标签方法; Transformer 编码器

中图分类号: TP181

文献标志码: A

Transformer-Based Image Classification Method of Semi-Supervised Partial Multi-Label

LI Yan^{1,2}, ZHOU Zhilong³, ZHU Biao⁴

(1. Faculty of Arts and Sciences, Beijing Normal University, Zhuhai 519085, China; 2. School of Applied Mathematics, Beijing Normal University at Zhuhai, Zhuhai 519087, China; 3. School of Statistics, Beijing Normal University, Beijing 100875, China;
4. College of Mathematics and Information Science, Hebei University, Baoding 071002, China)

Abstract: In SSPML (Semi-Supervised Partial Multi-Label Learning), complex correlations exist between the feature and label, and within them. Models leveraging label correlations are susceptible to interference from noisy labels in candidate label sets. T-SSPML (Transformer Encoder-Based Semi-Supervised Partial Multi-Label Learning), a novel semi-supervised partial multi-label image classification algorithm is proposed based on Transformer. A backbone network is constructed using Transformer encoders, taking both image features and label features as joint inputs, and a new loss function is defined. The multi-head self-attention mechanism is employed to capture holistic structural information and label interdependencies. A label masking mechanism is introduced to enhance the model's capability in learning label correlation information. Extensive experiments on multiple datasets demonstrate that the T-SSPML algorithm achieves superior classification performance compared to other representative methods in most cases.

Key words: partial multi-label learning; semi-supervised learning; image classification; consistency regularization; pseudo-labeling methods; Transformer encoder

收稿日期: 2025-08-29

基金项目: 国家自然科学基金资助项目(61976141)

作者简介: 李艳(1976—), 女, 河北衡水人, 北京师范大学教授, 博士, 硕士生导师, 主要从事机器学习、不确定信息处理等研究, (Tel)86-13733220291 (E-mail)ly@bnu.edu.cn。

0 引言

多标签学习^[1]旨在解决一个样本带有多个标签的分类问题。由于多标签分类是一个很常见问题,例如一张图片中含有多个物体等信息目标,则这张图片就拥有多个类别标签。因此近年来其在图像标注^[2]、文本分类^[3-4]和生物信息学^[5]等领域得到越来越多的研究和应用。传统的多标签学习通常假设每个样本都被正确标注,即样本的多个标签都是真实的。但在实际问题分类中由于领域知识或信息获取不充分,在对多标签样本进行类别标注时,难免将错误的类别(也称为噪声标签)加入到已包含正确类别的标签集中。这种样本的标签集包含所有正确标签同时存在错误标签的多标签学习称为偏多标签学习^[6],隶属于弱监督学习框架。

由于实际分类问题中样本的标注代价较高,而无标记样本大量存在,因此在很多应用领域所获取的数据集中可能只有部分带有标记的样本。并且其也存在于偏多标签学习中,即一部分数据是包含噪声标签的偏多标签样本;另一部分是大量无标签样本的学习场景,称为半监督偏多标签学习^[7]。解决这类问题的关键在于如何同时挖掘利用数据中的弱监督和无监督信息,实现对样本的正确分类。相比传统的半监督学习和偏多标签学习,半监督偏多标签学习所具备的监督信息更少,学习问题更有难度,因此建立具有良好性能的模型更具挑战性。目前单独的半监督或偏多标签学习的研究报道较多,而针对半监督偏多标签学习方面的研究报道还较为缺乏。

2020年Xie等^[7]首次提出半监督偏多标签学习(SSPML; Semi-Supervised Partial Multi-Label Learning),把真实标签作为潜变量形成低维嵌入空间,建立一种保持样本特征和标签空间相似性的半监督偏多标签学习方法。Guan等^[8]独立引入类似的半监督偏多标签分类问题(SPMLC; Semi-supervised Partial Multi-Label Classification),通过低秩和流形约束捕获样本的相邻结构将监督信息扩展到相邻实例上,在训练预测模型的同时估计潜在真实标签。但上述工作只考虑样本在特征空间的相似性,而没有在标签层面提取样本的相关性信息,更容易受噪声标签的干扰。2022年Tan等^[9]通过Hilbert-Schmidt独立性准则从特征和标签层面学习样本相关性,同时提出的框架以一种互利的方式实现样例特征映射的训练和真实标签的恢复。这些半监督偏多标签学习方法同时利用不精确标记和无标记样本的特征结构与标签的相关性,在人工特征的数据集上取得了更好的预测性能。但对样本的相关性表示是在特征和标签空间中分别进行,然后再学习两者间的一致性,没有同时融合其间及各自内部的复杂相关信息。而在SSPML学习问题中,不仅特征、标签、特征和标签间都存在相关性,且由于可能存在多个未知的错误标签,其相关性更加难以表示和学习。另外,由于偏多标签数据的候选标签集中噪声标签的存在,直接利用候选标签集得到的标签相关性信息不够可靠,会影响算法的分类性能。因此建立一个统一的学习框架,融合学习特征、标签以及其间的关联信息,在监督信息明显不足的情况下,有利于有效降低噪声标签的影响。

为此笔者针对含有大量无标签样本而只有少量偏多标签样本的图像数据,提出一种基于Transformer编码器、伪标签和一致性正则化方法的半监督偏多标签图像分类框架。将样例的标签与图像特征一同作为样例的特征输入,利用Transformer的多头注意力机制以及标签掩码机制学习输入特征(包含标签信息)间复杂的关联信息,同时在模型迭代中不断更新偏多标签样例的消歧结果。其将特征和标签信息共同融合并学习其间的关联性,协同消歧过程,放到统一的框架下迭代优化。定义包含3种监督信息的损失函数,可以在学习过程中同时利用监督、弱监督和无监督信息,更充分地利用数据的标签和结构信息。本方法适合直接处理图像大数据,降低噪声标签的影响,在含有大量无标签数据和少量偏多标签数据的复杂学习场景下提升分类模型的性能。

1 相关工作

半监督偏多标签学习是基于偏多标签学习和半监督多标签学习基础上构建的一种新的学习框架。

偏多标签学习(PML; Partial Multi-Label Learning)通过恢复样本的候选标签集合的潜在标签置信度,并对每个样本的标签进行重新排序进行预测^[6]。2019年Fang^[10]提出一种偏多标签学习方法 PARTICLE

(Partial Multi-Label Learning via Credible Label Elicitation), 通过相邻实例传播标签信息获取样本的高置信度标签, 并将其视为可信标签再进行训练, 提高标签传播的可靠性。而为克服现有方法中标签的共现概率邻接矩阵受限于训练数据集的问题, 2023 年 Yuan 等^[11] 针对图像分类提出一种基于图注意力 Transformer 网络, 可以学习标签间更丰富的语义信息。但目前多数 PML 方法仍未充分挖掘复杂的标签相关性。2024 年 Wang 等^[12] 利用 Transformer 块和条件层归一化提取高阶标签相关性, 可应用于不同领域含大量噪声标签的数据。虽然这些偏多标签学习方法有很好的效果, 可以处理含有噪声标签的多标签学习问题, 但不适用于含有大量无标签样本的半监督学习场景。

由于大量无标记数据的存在, 近年来半监督多标签学习也得到广泛研究, 用于处理含有无标签样本的多标签学习问题。最初的半监督多标签学习方法基本都是利用相邻样本将监督学习从有标签的样本传播给无标签的样本, 以此将无标签样本加入到训练中^[13-14]。由于多标签样本的标签空间有丰富的语义信息, 并且标签间存在关联, 因此单独对样本的特征空间学习并不能取得很好的分类效果, 于是标签空间信息的获取得到了人们重视。Lanchantin 等^[15] 通过将样本标签转换为 embedding 向量联合样本特征一同输入 Transformer 编码器中进行训练, 学习样本特征和标签间复杂的依赖关系。然而半监督多标签学习依赖于精确的标注, 不能直接处理有噪声标签的半监督偏多标签学习问题。

在半监督偏多标签学习的复杂学习场景中, 由于偏多标签样本的候选标签集中含有噪声标签, 真实标签的相关性被污染。Xie 等^[7] 提出的半监督偏多标签算法 SSPML, 为每个样例引入一个潜在的标签变量作为特征空间的低维嵌入, 通过保持特征空间和标签空间的相似性恢复真实标签, 同时通过标签空间的反馈自适应地更新相似性, 显著提高偏多标签分类模型在半监督场景下的性能。Guan 等^[8] 提出一种带有低秩和流形约束的半监督偏多分类模型, 通过学习局部标签的相关性和样本的近邻信息预测真实标签。Tan 等^[9] 通过引入 Hilbert-Schmidt 独立性准则导出标签相关性矩阵, 再通过恢复偏多标签样例和无标签样例的真实标签后更新相关性矩阵, 互利迭代训练分类器。上述方法本质上是学习特征空间和标签空间的一致性, 同时考虑样本在标签空间的相关性提高预测性能。但在特征和标签空间存在稀疏和噪声的情况下, 如何学习输入到输出的嵌入结构仍然是具有挑战性的问题。2024 年 Tan 等^[16] 提出一种半监督环境下基于紧凑共享标签嵌入的 PML 学习框架, 通过最小化从共享子空间到标签空间的重构损失, 同时最大化与特征空间的依赖关系, 将标签信息与输入判别信息链接到一个共享标签子空间中, 实现协同训练。但该方法有两个潜在假设, 一个是常见的平滑性假设, 即特征相似的样本具有相似的标签; 另外还假设噪声标签是完全随机产生的, 可能不符合某些现实情况。

由于大部分 SSPML 相关工作都是处理具有人工特征的数据集, 不能获取图像的结构信息, 且样本规模较小, 而半监督偏多标签图像数据一般具有规模大、维数高的特点, 特征和包含噪声的多个标签间存在更复杂的相关性。笔者从文献[15-16]中得到启发, 在这类学习问题中把特征和标签信息作为样本的整体输入, 不对噪声标签的分布做任何假设, 获取特征和标签空间的整体性结构信息, 从而降低候选标签集中噪声标签的干扰。

2 基于 Transformer 和一致正则化的半监督偏多标签分类方法

Transformer 在很多研究中已经展现强大的学习性能, 其能捕捉到输入特征间复杂而丰富的依赖关系, 从而保持图像和标签特征空间的相似性, 以及不同标签间的依赖关系; 一致性正则化机制是一种常用的半监督学习方法, 其要求模型对于同一样例的不同增强版本有相似的标签预测分布, 并在模型训练中利用无监督样本中所包含的信息。

2.1 问题陈述和方法框架

笔者的研究背景为训练数据集中含有极少量的确切标签样本 $\mathcal{D}_s$ 、少量的偏多标签样本 $\mathcal{D}_p$ 和大量的无标签样本 $\mathcal{D}_u$, 旨在提出一种新的半监督偏多标签学习框架。问题符号表示如下。

在 q 类的半监督偏多标签学习问题中, 偏多标签数据集 $\mathcal{D}_p = \{(x_{p_1}, S_{p_1}), \dots, (x_{p_m}, S_{p_m})\}$, 其中 $x_{p_i} \in X_p$, X_p 为偏多标签数据的样本空间; x_{p_i} 所对应的候选标签集 $S_{p_i} \subseteq Y$, $Y = \{1, 2, \dots, q\}$; 候选标签集包含样本的所有正确标签, 同时存在冗余错误标签。数据集中另外含有大部分无标签数据 $\mathcal{D}_u$, 其中 $\mathcal{D}_u = \{x_{u_1},$

$\cdots, x_{u_n}\}$, $x_{u_i} \in X_u$, X_u 为无标签数据的样本空间。学习任务是在数据集 $D = \mathcal{D}_s \cup \mathcal{D}_p \cup \mathcal{D}_u$ 上建立有效的分类器, 对未知类别的图像进行分类预测。其中 $|\mathcal{D}_s| \ll |\mathcal{D}_p \cup \mathcal{D}_u|$, 且 $|\mathcal{D}_p| < |D_u|$, $|A|$ 为集合 A 的基数。

笔者方法框架如图 1 和图 2 所示。其主要思想为通过伪标签方法和一致正则化方法同时利用确切标签样本的有监督信息、偏多标签样本的弱监督信息和无标签样本的无监督信息进行模型训练, 同时引入 Transformer 的多头注意力机制^[17]学习图像和标签特征空间的一致性, 以及不同标签间的依赖关系。

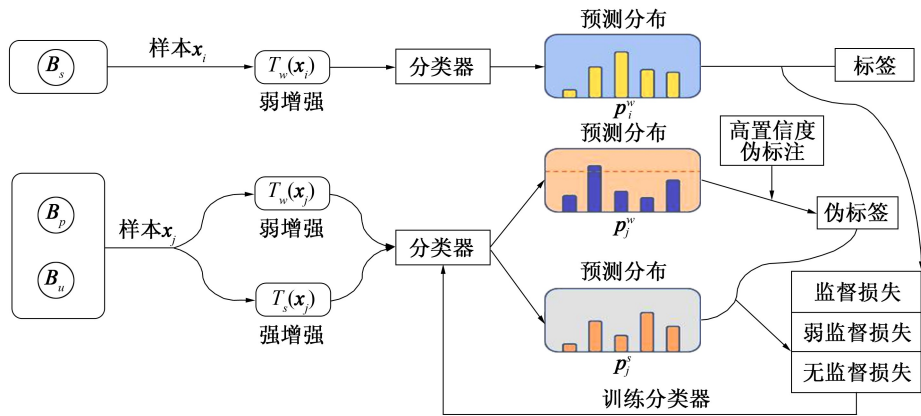


图 1 方法的整体框架
Fig. 1 Framework of the method

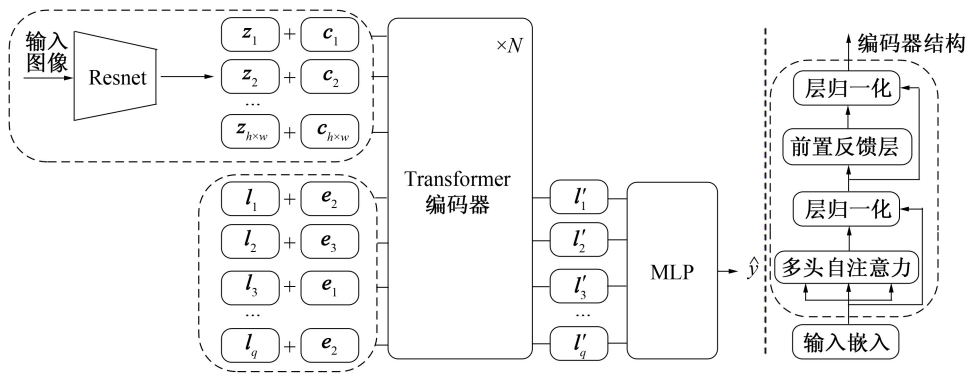


图 2 分类器框架图
Fig. 2 Classifier framework diagram

在图 1 中, B_s 、 B_p 和 B_u 分别为训练过程一个 Batch 中的有标签、偏多标签和无标签样本集合; $T_w(x_i)$ 和 $T_s(x_j)$ 分别为两个样本的弱和强增强变体; p_i^w 和 p_j^s 分别为两种变体的类别预测分布。训练过程中对一个 Batch 中的样本进行强弱数据增强, 输入到分类器模型分别进行预测, 得到预测分布。通过伪标签方法, 对偏多标签 B_p 和无标签样本 B_u 的弱增强样本输出进行标注, 并参与训练过程。通过一致性正则化方法计算出弱增强样本的伪标签和强增强样本输出的标准交叉熵损失, 作为在偏多标签训练集上的弱监督损失项 $\mathcal{L}_p$ 和无标签样本集上的无监督损失项 $\mathcal{L}_u$, 同确切标记样本 B_s 上得到的监督损失 $\mathcal{L}_s$ 一同构成最终的损失函数, 更新模型参数。随后利用更新的模型再次得到弱增强样本的伪标签, 依次循环迭代直到收敛, 最终得到训练好的分类模型。

图 1 的分类器框架如图 2 所示。笔者使用预训练好的 Resnet 作为特征提取器, 通过特征提取网络将每张输入样例图片 $x \in \mathbb{R}^{H \times W \times 3}$ 提取为 $Z \in \mathbb{R}^{h \times w \times d}$ 的特征张量, 其中 h, w 和 d 分别为高、宽和特征维度, 则输入样例图片可以看作提取分解成 $h \times w$ 个部分的特征向量, 同时将这些部分的位置信息通过 embedding 层提取为位置信息嵌入向量 $c_i \in \mathbb{R}^d, i \in \{1, \dots, h \times w\}$, 将每部分的特征向量与位置嵌入向量相加得到样例图片的特征输入。

每个标签通过 embedding 层提取为与图片特征相同维度的 label embedding 向量 $L \in \mathbb{R}^{q \times d}$, 标签的真、

假和未知3种标签状态同样通过 embedding 层提取为3个不同的 state embedding 向量 $E = \{e_1, e_2, e_3\} \in \mathbb{R}^{3 \times d}$ 并将其嵌入 label embedding 向量中,然后将图片和标签特征一同输入 Transformer 编码器,最后编码器的输出通过 MLP(Multi-Layer Perceptron)层得到样本的标签预测分布 $\hat{y}$ 。

Transformer 编码器的输入为

$$G = \{\bar{z}_i, \bar{l}_j | \bar{z}_i = z_i + c_i, \quad i \in \{1, \dots, h \times w\}, \quad \bar{l}_j = l_j + e_j, \quad j \in \{1, \dots, q\}\}.$$

Transformer 编码器通过自注意力机制学习 $g_i \in G$ 和 $g_j \in G$ 间的注意力权重 α_{ij} , 通过 α_{ij} 更新特征向量 g_i 得到下一层的输入向量 g'_i 。如下:

$$\alpha_{ij} = \text{softmax}((w^q g_i)^T (w^k g_j) / \sqrt{d}), \quad (1)$$

$$\bar{g}_i = g_i + \sum_{j=1}^M \alpha_{ij} w^v g_j, \quad (2)$$

$$g'_i = w_2 \text{ReLU}(w_1 \bar{g}_i + b_1) + b_2, \quad (3)$$

其中 w^q 为 query 权重矩阵, w^k 为 key 权重矩阵, w^v 为 value 权重矩阵, w_1 和 w_2 为两个全连接层的变换矩阵, $\{w^q, w^k, w^v, w_1, w_2\} \in \mathbb{R}^{d \times d}$ 。最终得到层输出 $G' = \{z'_i, l'_j | i \in \{1, \dots, h \times w\}, j \in \{1, \dots, q\}\}$ 作为下一层的输入,依次重复更新。

2.2 一致性正则化和伪标签方法

一致性正则化^[18]是半监督算法中常用的一种方法,其主要思想为一个样本的不同扰动变体的模型预测输出应当相似。样本的扰动通常采用数据增强的方式,扰动同时可以实现数据集的扩充,提高模型的泛化性能。对不同变体的模型输出间的差异,可以基于差异最小化进行模型训练。目标函数如下:

$$\sum_{i=1}^M \|p(y|T(x_i)) - p(y|x_i)\|_2^2, \quad (4)$$

其中 M 为所考虑样本的总数, x_i 为第 i 个样本, $T(x_i)$ 为样本 x_i 的增强样本,得到的增强样本的类别预测分布是 p_i^T 。需要指出,因为增强函数 $T(\cdot)$ 的随机性,式(4)中的两个部分值并不相同。且笔者的学习框架中增强函数 $T(\cdot)$ 和损失函数都可以根据问题进行改变替换,并不局限于特定函数。如图1所示,笔者采用了弱和强增强两种方法,弱增强函数为 $T_w(\cdot)$,强增强函数为 $T_s(\cdot)$ 。在实验部分,在弱增强方法中只对图像进行了随机的翻转和平移增强策略;在强增强方法中,使用了 RandAugment^[19] 数据增强策略,其是 AutoAugment^[20] 的一种变体,不需要使用标记数据提前学习,对每个样本都进行随机增强。

所提方法同时采用半监督学习中典型的伪标签技术^[21](见图1),通过设定置信度阈值,将高于阈值的样本标签标注为真,低于阈值的样本标签标注为假,得到样本标签的 one-hot 向量,并将偏多标签样本候选标签集中的错误标签剔除,同时无标签样本也得到标注。

2.3 损失函数

在笔者所描述的问题背景下,数据集只含有极少量的确切标签样本和少量偏多标签样本,以及大量的无标签样本。笔者方法的损失函数由3部分组成,其中 $\mathcal{L}_s$ 为确切监督损失项:

$$\mathcal{L}_s = \frac{1}{B_s} \sum_{i=1}^{B_s} H(p_i, p_i^w), \quad (5)$$

$\mathcal{L}_p$ 为偏多标签监督损失项, $\mathcal{L}_u$ 为无监督损失项。确切标签数据只通过弱增强处理产生弱增强样本 $T_w(x_i)$, $\mathcal{L}_s$ 为 $T_w(x_i)$ 的模型输出 p_i^w 与其确切标签 p_i 的标准交叉熵损失; B_s 为一个 Batch 中确切标签样本个数。

对偏多标签样本,通过弱增强处理产生弱增强样本 $T_w(x_i)$,将 $T_w(x_i)$ 的模型输出 p_i^w 通过伪标签方法转换为 one-hot 向量 $\hat{p}_i^w$,其中伪标签置信度阈值设置为 0.5。由于样本的正确标签都存在于候选标签集中,为保证损失的可靠性,只有伪标签方法标注为真(值为1)的标签中至少一半是存在于候选标签集时,这个偏多标签样本的损失才计入损失,同时将不存在于候选标签集且标注为真的伪标签置为假(值为0),则保证了标注为真的伪标签一定存在于原始的标签候选集, $\hat{p}_i^w$ 也作为此样本下次训练的标签输入,依次迭代更新;通过强增强处理产生强增强样本 $T_s(x_i)$, $\mathcal{L}_p$ 为 $T_w(x_i)$ 的硬输出 $\hat{p}_i^w$ 与

$T_s(\mathbf{x}_i)$ 的预测分布 $\mathbf{p}_i^s$ 间的标准交叉熵损失。用 B_p 表示一个 Batch 中偏标记样本个数, $\mathcal{L}_p$ 如下:

$$\mathcal{L}_p = \frac{1}{B_p} \sum_{i=1}^{B_p} H(\hat{\mathbf{p}}_i^w, \mathbf{p}_i^s). \quad (6)$$

对无标记样本的处理方法和偏多标签样本的处理方法类似, 分别用 $T_w(\cdot)$ 和 $T_s(\cdot)$ 生成弱增强样本 $T_w(\mathbf{x}_i)$ 和强增强样本 $T_s(\mathbf{x}_i)$, 将 $T_w(\mathbf{x}_i)$ 的模型输出 $\mathbf{p}_i^w$ 通过伪标签方法转换为 one-hot 向量 $\hat{\mathbf{p}}_i^w$, $\mathcal{L}_u$ 为 $T_s(\mathbf{x}_i)$ 的模型输出 $\mathbf{p}_i^s$ 和 $\hat{\mathbf{p}}_i^w$ 的标准交叉熵损失。 B_u 表示一个 Batch 中无标记样本个数, $\mathcal{L}_u$ 如下:

$$\mathcal{L}_u = \frac{1}{B_u} \sum_{i=1}^{B_u} H(\hat{\mathbf{p}}_i^w, \mathbf{p}_i^s). \quad (7)$$

由于无标签样本没有确切和候选标签, 为衡量无标签样本伪标签 $\hat{\mathbf{p}}_i^w$ 的可靠性, 根据偏多标签样本的训练状态设置了无标签样本损失权重 μ , 如下:

$$\mu = \frac{N_t}{D_p}, \quad (8)$$

其中 N_t 为 t 阶段时标注了伪标签的偏多标签样本数, D_p 为偏多标签样本的总数。随着训练阶段的深入, N_t 也逐渐增大, 无标签样本的伪标签也更加可靠, 损失权重 μ 也随之增大。

最终, 总的损失函数 $\mathcal{L}_{\text{all}}$ 为 3 部分损失项的加权和:

$$\mathcal{L}_{\text{all}} = \mathcal{L}_s + \mathcal{L}_p + \mu \mathcal{L}_u. \quad (9)$$

2.4 标签消歧和标签掩码

由于偏多标签样本的候选标签集中存在噪声标签, 直接用候选标签集构造标签状态向量更新公式 $\bar{\mathbf{l}}_j = \mathcal{L}_j + \mathbf{e}_j$, $j \in \{1, \dots, q\}$, 可能会对模型的分类性能产生负面影响。因此, 如模型框架图 1 所示, 类似半监督偏标签学习方法, T-SSPML (Transformer-based Semi-Supervised Partial Multi-label Learning) 在训练过程中通过伪标签方法, 对偏多标签样本的弱增强变体 x_i 的输出结果 $\mathbf{p}_i^w$ 进行消歧, 降低噪声标签的负面影响。如下:

$$\hat{\mathbf{p}}_i^w[k] = \begin{cases} 1, & \text{if } \mathbf{p}_i^w[k] \geq 0.5 \text{ and } k \in S_i, \\ 0, & \text{其他,} \end{cases} \quad (10)$$

其中 $\mathbf{p}_i^w[k]$ 为类别 k 的预测值, S_i 为偏多标签样本的候选标签集。当类别 k 属于候选标签集 S_i 且预测值不小于 0.5 时, 将类别 k 标注为样本 $\mathbf{x}_i$ 的真实标签, 其余标签都标注为假标签。利用消歧结果 $\hat{\mathbf{p}}_i^w$ 构造标签状态向量, 如下:

$$\mathbf{e}_k^i = \begin{cases} \mathbf{e}_1, & \text{if } \hat{\mathbf{p}}_i^w[k] = 1, \\ \mathbf{e}_2, & \text{if } \hat{\mathbf{p}}_i^w[k] = 0, \\ \mathbf{e}_3, & \text{if } \hat{\mathbf{p}}_i^w[k] = -1, \end{cases} \quad (11)$$

其中 $\mathbf{e}_k^i$ 为样本 x_i 的第 k 个类别标签状态向量。

在训练的最初阶段, 偏多标签数据的消歧结果 $\hat{\mathbf{p}}_i^w$ 初始化公式如下:

$$\hat{\mathbf{p}}_i^w[k] = \begin{cases} -1, & \text{if } k \in S_i, \\ 0, & \text{其他,} \end{cases} \quad (12)$$

其中候选标签集中标签的消歧结果初始化为 -1, 表示标签状态为未知状态, 候选标签集之外标签的消歧结果置为 0, 表示标签状态为假。无标签样本的所有标签状态始终置为未知状态。

笔者进一步在分类器训练过程中采用标签掩码机制, 旨在让模型能通过部分真实标签信息, 推断出另一部分被遮掩的真实标签的存在, 从而使模型在学习标签依赖等相关性信息方面有更强的能力。对有标签样本的弱增强变体和偏多标签样本的强增强变体, 在式(11)生成标签状态向量时, 采用标签掩码机制随机遮掩 1/4 的真实标签状态为未知状态:

$$\hat{\mathbf{e}}^i = \text{mask}(\mathbf{e}^i), \quad (13)$$

其中 $\text{mask}(\cdot)$ 为标签掩码函数, $\mathbf{e}^i$ 为样本 $\mathbf{x}_i$ 的标签状态向量, $\hat{\mathbf{e}}^i$ 为样本 $\mathbf{x}_i$ 进行标签掩码后的标签状态向量。

3 实验设置和结果

3.1 数据集介绍和实验设置

如表1所示,实验选取4个多标签图像分类数据集:COCO(Common Objects in COntext)^[22]、Pascal VOC2007^[23]、mirflickr25k^[24]和CelebA^[25],其图像样本数从9 963到202 599;多标签数量从20个到80个。通过不同设置生成相应的偏多标签数据集,使每个数据集中包含极少量精确标签样本、较少偏多标记样本和大量无标记样本。其中设置在 q 分类数据集中仅含有 $8q$ 个确切标签样本,即平均每个类别仅有8个正确标注的样本,共生成12个不同设置的数据集。

表1 数据集
Tab.1 Dataset overview

数据集	样本数	类别数	平均标签数	数据集	样本数	类别数	平均标签数
COCO	82 738	80	2.90	mirflickr25k	25 000	38	4.72
Pascal VOC2007	9 963	20	2.47	CelebA	202 599	40	9.03

对 q 分类问题中的每个偏多标签样本,除满足样本的真实标签一定存在于候选标签集合中,以真实标签个数的比例 $\theta \in \{50\%,100\%,150\%\}$ ^[9]随机添加噪声标签到候选标签集生成偏多标签数据。即 θ 是噪声标签与真实标签的比值, θ 越大表示添加的噪声标签越多。此外,为研究不同偏多标签样本数量和无标签样本数量的比例对于分类器性能的影响,还设置不同的偏多标签数据 $\mathcal{D}_p$ 与无标签数据 $\mathcal{D}_u$ 所含样本数量的比例,具体的比例大小 $|\mathcal{D}_p|/|\mathcal{D}_u| \in \{1/9,1/4,2/3\}$,即带有偏标记的样本分别为10%,20%,40%,最多占整个数据集中样本的40%。

笔者方法和5个相关的代表性方法进行了比较:

- 1) PML-NI(Partial Multi-Label Learning with Noisy label Identification)^[6],一种偏多标签学习算法,能识别由实例内容不明确引起的噪声标签;
- 2) SSPML^[7],一种半监督偏多标签学习算法,引入了一种潜在的标签变量作为特征空间的低维嵌入,在标签变量的监督下对分类器进行训练;
- 3) PARTICLE-MAP(PARTIal multi-label learning via Credible Label Elicitation)^[10],一种偏多标签学习算法,通过标签传播标注无标签样本进行迭代训练;
- 4) PML-LD(Partial Multi-Label Learning with Label Distribution)^[26],一种偏多标签学习算法,利用特征空间的拓扑信息和标签之间的相关性恢复样本的标签分布;
- 5) TRAM(TRASductive Multi-label classification)^[27],一种多标签学习算法,其将多标签学习转换为一种估计标签概念组成的优化问题,并提出了一种标注无标签样本的算法,可同时利用有标签样本和无标签样本进行分类器训练。

笔者采用在ImageNet数据集上预训练好的ResNet-101^[28]作为图片特征提取网络,训练开始时图片大小调整为640×640像素,在进行强或弱增强后随机裁剪为576×576像素,再经过ResNet-101网络后提取大小为18×18×2 048的特征张量,为公平比较,同样使用ResNet-101提取2 048维度大小的图片特征作为5个对比算法的特征输入;Transformer编码器由3层Transformer组成,采用4个注意力头,不同层间的参数不共享。在训练开始阶段,偏多标签样本的候选标签的标签状态都置为未知,其余标签的标签状态置为假,无标签样本的标签状态在训练过程中一直置为未知。

笔者选取4种常用的多标签分类性能评价指标^[1]:平均精度(average precision)、排序损失(ranking loss)、单误差(one-error loss)、汉明损失(hamming loss)。其中平均精度指标的值越大代表分类器的性能越好,排序损失、单误差和汉明损失3个指标的值越小代表分类器的性能越好。取均值±标准差作为最终结果,如表2~表5所示。其中“T-SSPML”表示笔者所提方法,“·黑体数字”表示每列最优结果;“*数字”代表每列次优结果。

表 2 各算法在 COCO 数据集上的结果(mean±std)
Tab. 2 Results of algorithms on COCO dataset(mean±std)

$ \mathcal{D}_p / \mathcal{D}_u =1:9$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.569±0.048	* 0.123±0.014	• 0.286±0.024	* 0.029±0.014
PML-NI	0.531±0.047	0.146±0.008	0.345±0.031	0.037±0.019
PAR-MAP	0.336±0.062	0.314±0.034	0.425±0.048	0.034±0.017
PML-LD	0.378±0.037	0.317±0.026	0.432±0.035	0.106±0.010
SSPML	0.439±0.052	0.195±0.023	0.435±0.021	0.038±0.016
TRAM	* 0.544±0.046	• 0.114±0.018	* 0.303±0.032	• 0.018±0.013
$ \mathcal{D}_p / \mathcal{D}_u =1:4$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.583±0.041	• 0.074±0.018	• 0.237±0.019	* 0.025±0.012
PML-NI	0.551±0.037	0.118±0.012	* 0.282±0.027	0.033±0.007
PAR-MAP	0.375±0.068	0.285±0.043	0.392±0.036	0.036±0.023
PML-LD	0.359±0.046	0.314±0.024	0.414±0.029	0.097±0.016
SSPML	0.493±0.052	0.187±0.017	0.403±0.018	0.036±0.026
TRAM	* 0.574±0.043	* 0.080±0.015	0.305±0.023	• 0.019±0.008
$ \mathcal{D}_p / \mathcal{D}_u =2:3$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.628±0.028	• .065±0.011	• 0.221±0.014	* 0.027±0.016
PML-NI	* 0.601±0.031	0.131±0.015	* 0.268±0.021	0.032±0.014
PAR-MAP	0.417±0.054	0.270±0.032	0.372±0.025	0.034±0.026
PML-LD	0.396±0.045	0.316±0.021	0.381±0.033	0.067±0.023
SSPML	0.543±0.037	0.147±0.013	0.316±0.017	0.035±0.015
TRAM	0.592±0.023	* 0.068±0.008	0.275±0.013	• 0.017±0.010

表 3 各算法在 Pascal 数据集上的结果(mean±std)
Tab. 3 Results of algorithms on Pascal dataset(mean±std)

$ \mathcal{D}_p / \mathcal{D}_u =1:9$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.642±0.073	• 0.121±0.016	• 0.436±0.021	* 0.068±0.008
PML-NI	* 0.621±0.094	* 0.157±0.021	* 0.445±0.018	0.210±0.025
PAR-MAP	0.443±0.124	0.241±0.038	0.662±0.040	0.091±0.014
PML-LD	0.394±0.072	0.316±0.034	0.608±0.072	0.129±0.018
SSPML	0.438±0.111	0.292±0.026	0.582±0.051	0.079±0.007
TRAM	0.573±0.059	0.185±0.013	0.476±0.025	0.057±0.008
$ \mathcal{D}_p / \mathcal{D}_u =1:4$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.675±0.054	• 0.112±0.013	• 0.421±0.029	0.065±0.012
PML-NI	0.632±0.072	0.146±0.010	* 0.427±0.015	* 0.061±0.009
PAR-MAP	0.457±0.105	0.271±0.026	0.593±0.036	0.087±0.014
PML-LD	0.425±0.092	0.328±0.025	0.573±0.058	0.125±0.025
SSPML	0.526±0.083	0.239±0.019	0.534±0.042	0.072±0.017
TRAM	* 0.635±0.056	* 0.125±0.016	0.453±0.014	• 0.053±0.006
$ \mathcal{D}_n / \mathcal{D}_u =2:3$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.714±0.046	• 0.103±0.012	• 0.412±0.020	* 0.056±0.013
PML-NI	0.652±0.053	0.173±0.015	* 0.425±0.022	0.059±0.008

(续表 3)

$ \mathcal{D}_n / \mathcal{D}_u =2:3$				
算法	平均精度	排序损失	单误差	汉明损失
PAR-MAP	0.479±0.087	0.228±0.024	0.583±0.042	0.086±0.023
PML-LD	0.412±0.091	0.302±0.032	0.553±0.036	0.111±0.027
SSPML	0.545±0.063	0.198±0.021	0.525±0.028	0.075±0.014
TRAM	* 0.691±0.051	* 0.116±0.016	0.431±0.013	• 0.051±0.005

表 4 各算法在 mirflickr25k 数据集上的结果 (mean±std)

Tab. 4 Results of algorithms on mirflickr25k dataset (mean±std)

$ \mathcal{D}_p / \mathcal{D}_u =1:9$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.516±0.054	• 0.218±0.021	• 0.470±0.035	* 0.133±0.013
PML-NI	0.443±0.043	0.263±0.024	0.512±0.046	0.153±0.018
PAR-MAP	0.472±0.081	0.246±0.032	0.571±0.031	0.142±0.023
PML-LD	0.480±0.062	0.243±0.027	0.563±0.083	0.146±0.019
SSPML	* 0.501±0.047	* 0.237±0.018	* 0.487±0.042	• 0.122±0.011
TRAM	0.429±0.034	0.286±0.036	0.593±0.053	0.203±0.025

$ \mathcal{D}_n / \mathcal{D}_u =1:4$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.541±0.042	• 0.203±0.017	• 0.427±0.043	* 0.132±0.016
PML-NI	0.474±0.044	0.232±0.023	0.442±0.047	0.152±0.015
PAR-MAP	0.516±0.068	* 0.224±0.025	0.523±0.029	0.136±0.008
PML-LD	0.488±0.053	0.236±0.030	0.536±0.074	0.137±0.014
SSPML	* 0.521±0.041	0.229±0.023	* 0.431±0.037	• 0.112±0.007
TRAM	0.463±0.025	0.267±0.029	0.553±0.044	0.184±0.019

$ \mathcal{D}_n / \mathcal{D}_u =2:3$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.577±0.049	• 0.191±0.024	• 0.413±0.031	* 0.128±0.008
PML-NI	0.531±0.037	0.233±0.034	0.434±0.026	0.145±0.021
PAR-MAP	0.564±0.055	0.206±0.022	0.503±0.054	0.129±0.013
PML-LD	0.551±0.045	0.216±0.025	0.488±0.036	0.132±0.015
SSPML	* 0.568±0.038	* 0.202±0.018	* 0.427±0.027	0.111±0.005
TRAM	0.511±0.022	0.254±0.023	0.518±0.050	0.155±0.014

表 5 各算法在 CelebA 数据集上的结果 (mean±std)

Tab. 5 Results of algorithms on CelebA dataset (mean±std)

$ \mathcal{D}_p / \mathcal{D}_u =1:9$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.564±0.038	• 0.181±0.023	* 0.157±0.043	* 0.172±0.016
PML-NI	0.472±0.062	0.315±0.045	0.442±0.069	0.269±0.032
PAR-MAP	0.524±0.026	0.256±0.032	0.185±0.052	0.185±0.027
PML-LD	0.545±0.043	0.208±0.019	0.162±0.038	0.186±0.019
SSPML	0.536±0.025	0.212±0.028	0.179±0.062	0.192±0.023
TRAM	* 0.548±0.035	* 0.193±0.026	• 0.154±0.037	• 0.165±0.028

$ \mathcal{D}_p / \mathcal{D}_u =1:4$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.598±0.031	• 0.174±0.016	• 0.141±0.032	* 0.168±0.017
PML-NI	0.519±0.043	0.276±0.038	0.358±0.057	0.267±0.029
PAR-MAP	0.534±0.032	0.195±0.024	0.169±0.045	0.177±0.024

(续表5)

$ \mathcal{D}_p / \mathcal{D}_u =1:4$				
算法	平均精度	排序损失	单误差	汉明损失
PML-LD	* 0.579±0.027	0.186±0.014	0.157±0.031	0.176±0.025
SSPML	0.552±0.045	0.193±0.027	0.169±0.046	0.182±0.026
TRAM	0.568±0.034	* 0.184±0.013	* 0.143±0.042	0.146±0.025

$ \mathcal{D}_p / \mathcal{D}_u =2:3$				
算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	• 0.653±0.025	• 0.152±0.012	• 0.114±0.036	* 0.153±0.012
PML-NI	0.610±0.037	0.224±0.023	0.205±0.048	0.246±0.024
PAR-MAP	0.587±0.031	0.191±0.019	0.143±0.035	0.164±0.018
PML-LD	* 0.652±0.029	0.170±0.017	0.132±0.028	0.154±0.021
SSPML	0.624±0.037	0.183±0.022	0.155±0.051	0.177±0.032
TRAM	0.631±0.021	* 0.163±0.013	* 0.130±0.025	• 0.138±0.018

3.2 结果分析

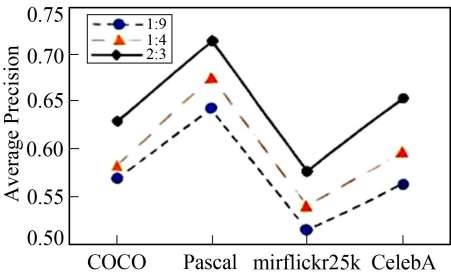
从所有算法在4个数据集以及不同的偏多标签和无标签数据比例下的实验结果如表2~表5所示,从表2~表5可得到如下结果。

1) 对整体结果,比较6种算法在4个数据集、3种标签比例、4种性能指标,共48种情况中的表现,所提出的T-SSPML方法展现了明显优势,取得34次最优、12次次优结果。尤其在平均精度、排序损失和单误差这3个评价指标下的36种情况中,笔者的算法几乎都取得了最佳的性能,只有2种情况取得次优性能,分别是表1中 $|\mathcal{D}_p|/|\mathcal{D}_u|=1:9$ 的排序损失值和表4中 $|\mathcal{D}_p|/|\mathcal{D}_u|=1:9$ 的单误差值,这可能是因为 $|\mathcal{D}_p|/|\mathcal{D}_u|$ 的值比较小时,偏多标签样本数量过少,无标签样本的伪标签可靠性不高,影响了部分指标的性能效果。

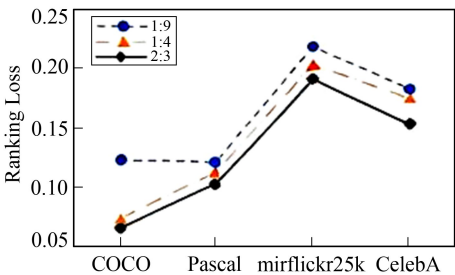
2) 对平均分类精度,所提方法全部取得最优结果,相比次优精度提高1%~7%。其次是TRAM算法,在48种情况中取得11个最优、15个次优结果;再次是SSPML算法,取得3个最优和8个次优结果,这是因为这两种算法都可以同时处理多标签和无标签数据,而其余的3种算法都是偏多标签学习方法,不适用于半监督偏多标签场景。

3) 对汉明损失评价指标,笔者的算法虽然没有取得最优性能,但几乎都取得次优,且与最优结果相当,仅相差1%左右。其原因可能是:与前3种评价指标计算方式不同,汉明损失是基于标签分类真假的布尔值(即标签值为0或1)而非标签置信度排序进行计算,而所提方法采用标签特征嵌入方法学习标签间的依赖关系,当一些标签为真时会增加与之相关的标签的分类值大小,因此有些假标签被分类为真标签,即可能会产生假阳性标签,一定程度降低了汉明损失的指标性能。

4) 偏多标签数据和无标签数据的比例参数对模型的分类结果也有重要影响,从图3在同一数据集上的结果可以看到,随着偏多标签数据所占据比例的增大,算法的性能指标值有明显的提升,说明偏多标签数据中的偏多标签信息得到了充分的发掘利用。



a 平均精度



b 排序损失

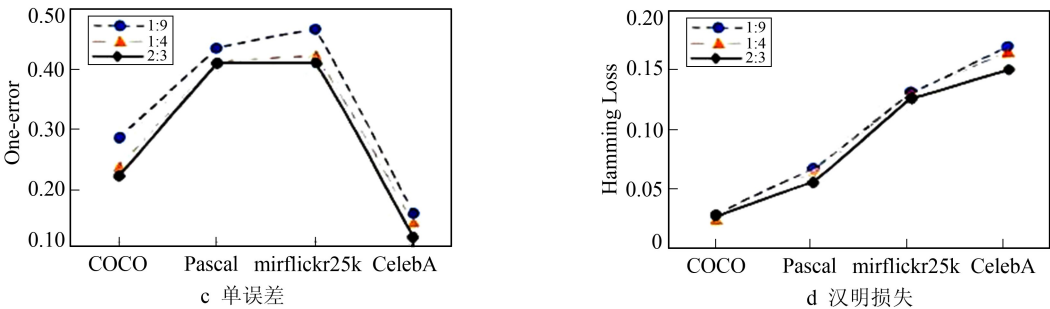


图 3 各性能指标随 $|\mathcal{D}_p|/|\mathcal{D}_u|$ 取值的变化趋势

Fig. 3 Trends of performance metrics with $|\mathcal{D}_p|/|\mathcal{D}_u|$ parameter variations

3.3 消融实验

为说明标签状态嵌入和标签掩码机制的有效性,在 COCO 数据集上 $|\mathcal{D}_p|/|\mathcal{D}_u|=2/3$ 的情况下进行了消融实验。首先移除标签状态嵌入,方法名表示为 Ours-1;然后移除标签掩码机制,即在运用公式 $\bar{l}_j = l_j + e_j, j \in \{1, \dots, q\}$ 生成样本标签特征向量时,将所有标签的状态嵌入向量置为未知状态向量 e_3 ,方法名表示为 Ours-2。

通过表 6 可看出, Ours-1 的分类性能显著高于 Ours-2 的分类性能,说明标签状态嵌入对模型性能提升较大;同时原始 T-SSPML 分类性能高于 Ours-1 分类性能,说明标签掩码机制对于模型性能也有一定的提升作用。

表 6 消融实验结果
Tab. 6 Ablation study results

算法	平均精度	排序损失	单误差	汉明损失
T-SSPML	0.628±0.028	0.065±0.011	0.221±0.014	0.027±0.016
Ours-1	0.612±0.026	0.083±0.015	0.239±0.031	0.028±0.018
Ours-2	0.587±0.031	0.115±0.013	0.253±0.028	0.032±0.021

4 结 语

笔者针对只含有少量偏标记样本和大量无标记样本的图像分类问题,提出一种新的半监督偏多标签学习框架。利用 Transformer 的自注意力和掩码机制学习图像特征和标签空间的相互关联信息,同时采用一致性正则化和伪标签方法,利用偏多标签和无标签数据中的弱和无监督信息进行模型训练。其特点在于将特征和标签同时作为输入,综合 3 种监督损失,把学习特征、标签及特征和标签间的复杂相关性融入统一的框架进行学习,降低噪声标签的影响,实现更精准的分类预测。在基于 4 个图像数据生成的 12 个不同情况的数据集上进行大量实验,采用多个评价指标对模型性能进行评估,所提方法相比其他对比算法在平均精度、排序损失、单误差评价指标上绝大多数情况下取得了更好的分类性能。尚存在的不足之处包括一些参数仍存在优化空间,比如损失函数中的比例权重 μ 及生成伪标签的阈值,未来可探讨其对结果的影响;另外将进一步拓展所提方法的应用范围,如考虑将本研究应用于多模态的半监督偏标记学习场景中。

参考文献:

[1] 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.

[2] LI Q, PENG X J, QIAO Y, et al. Learning Label Correlations for Multi-Label Image Recognition with Graph Networks [J]. Pattern Recognition Letters, 2020, 138: 378-384.

[3] LI X M, OUYANG J H, ZHOU X T. Centroid Prior Topic Model for Multi-Label Classification [J]. Pattern Recognition Letters, 2015, 62: 8-13.

[4] HE Z Y, LYU P, WU J. Minimum Classification Error Rate Training of Supervised Topic Mixture Model for Multi-Label Text

- Categorization [C] // 2014 9th International Symposium on Chinese Spoken Language Processing. Piscataway, NJ, USA: IEEE, 2014: 39-43.
- [5] 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.
- [6] XIE M K, HUANG S J. Partial Multi-Label Learning with Noisy Label Identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3676-3687.
- [7] XIE M K, HUANG S J. Semi-Supervised Partial Multi-Label Learning [C] // Industrial Conference on Data Mining. Piscataway, NJ, USA: IEEE Publisher, 2020: 691-700.
- [8] GUAN Y Y, ZHANG B X, LI W H, et al. Semi-Supervised Partial Multi-Label Classification with Low-Rank and Manifold Constraints [J]. Pattern Recognition Letters, 2021, 151: 112-119.
- [9] TAN A H, LIANG J Y, WU W Z, et al. Semi-Supervised Partial Multi-Label Classification via Consistency Learning [J/OL]. Pattern Recognition, 2022, 131(1): 108839 [2025-07-25]. <https://doi.org/10.1016/j.patcog.2022.108839>.
- [10] FANG J P, ZHANG M L. Partial Multi-Label Learning via Credible Label Elicitation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(10): 3518-3525.
- [11] YUAN J, CHEN S K, ZHANG Y, et al. Graph Attention Transformer Network for Multi-Label Image Classification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 19(4): 1-16.
- [12] WANG Z W, LIU F H, HAN M J, et al. PML-ED: A Method of Partial Multi-Label Learning by Using Encoder-Decoder Framework and Exploring Label Correlation [J/OL]. Information Sciences, 2024, 661: 120165 [2025-07-25]. <https://doi.org/10.1016/j.ins.2024.120165>.
- [13] 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, NY, USA: ACM, 2017: 1305-1314.
- [14] LI Q M, WU X M, LIU H, et al. Label Efficient Semi-Supervised Learning via Graph Filtering [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 9582-9591.
- [15] LANCHANTIN J, WANG T L, ORDONEZ V, et al. General Multi-Label Image Classification with Transformers [C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2021: 16478-16488.
- [16] TAN A H, WU W Z. Partial Multi-Label Learning via Semi-Supervised Subspace Collaboration [J/OL]. Knowledge-Based Systems, 2024, 287: 111444 [2025-07-25]. <https://doi.org/10.1016/j.knosys.2024.111444>.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C] // Advances in Neural Information Processing Systems. New York, NY, USA: Curran Associates, 2017: 5998-6008.
- [18] MIYATO T, MAEDA S I, KOYAMA M, et al. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1979-1993.
- [19] CUBUK E D, ZOPH B, SHLENS J, et al. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2020: 3008-3017.
- [20] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: Learning Augmentation Strategies from Data [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 113-123.
- [21] LEE D H. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks [C] // Workshop on Challenges in Representation Learning, ICML. New York, NY, USA: PMLR Publisher, 2013: 1-6.
- [22] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context [C] // European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 740-755.
- [23] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The Pascal Visual Object Classes Challenge 2007 [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [24] HUISKES M J, LEW M S. The MIR Flickr Retrieval Evaluation [C] // The 1st ACM International Conference on Multimedia Information Retrieval. New York, NY, USA: ACM Press, 2008: 39-43.
- [25] LIU Z W, LUO P, WANG X G, et al. Deep Learning Face Attributes in the Wild [C] // 2015 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2015: 3730-3738.
- [26] XU N, LIU Y P, GENG X. Partial Multi-Label Learning with Label Distribution [C] // AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI, 2020: 6510-6517.
- [27] 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.
- [28] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 770-778.