软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(2):749−761 [doi: 10.13328/j.cnki.jos.007447] [CSTR: 32375.14.jos.007447]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于不变性注入的多标记类属特征学习*

杭均一[1,2], 张敏灵[1,2]

[1](东南大学 计算机科学与工程学院, 江苏 南京 210096)

[2](计算机网络和信息集成教育部重点实验室 (东南大学), 江苏 南京 210096)

通信作者: 张敏灵, E-mail: zhangml@seu.edu.cn

摘　要: 类属特征是一种解决多标记分类问题的有效策略. 通过为不同标记的判别过程提供不同的定制特征, 类属特征能够同时兼顾各个标记潜在不同的判别偏好, 进而改善多标记分类模型的泛化性能. 为学习类属特征, 已有方法通常关注于利用特征处理技术对样本中标记判别的相关特征进行提取. 不同于上述常规做法, 尝试从特征不变性的视角解决类属特征的学习问题: 通过操纵标记判别的无关特征, 为分类模型注入关于无关特征的不变性, 从而充分地兼顾各个标记的判别偏好. 相应地, 提出一种基于不变性注入的多标记类属特征学习方法 INVA. INVA 方法通过估计特征协方差矩阵捕获各个标记的类内特征变化, 从而辨识标记判别的无关特征; 通过求解扰动风险最小化问题, 赋予分类模型关于无关特征变化的不变性. 进一步地, 推导扰动风险最小化问题的上界, 提高了方法的计算效率. 在多标记基准数据集上, 与已有方法进行全面的实验对比, 验证所提方法的有效性.

关键词: 多标记分类; 类属特征; 特征不变性

中图法分类号: TP18

中文引用格式: 杭均一, 张敏灵. 基于不变性注入的多标记类属特征学习. 软件学报, 2026, 37(2): 749–761. http://www.jos.org.cn/1000-9825/7447.htm

英文引用格式: Hang JY, Zhang ML. Multi-label Label-specific Feature Learning Based on Invariance Injection. Ruan Jian Xue Bao/Journal of Software, 2026, 37(2): 749–761 (in Chinese). http://www.jos.org.cn/1000-9825/7447.htm

Multi-label Label-specific Feature Learning Based on Invariance Injection

HANG Jun-Yi[1,2], ZHANG Min-Ling[1,2]

[1](School of Computer Science and Engineering, Southeast University, Nanjing 210096, China)

[2](Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China)

Abstract: Label-specific features serve as an effective strategy for addressing multi-label classification tasks. By tailoring discriminative features to the individual preferences of each label, such features enhance the generalization capability of classification models. Existing methods typically focus on manipulating features to extract those relevant to label discrimination. Rather than following this conventional approach, this study explores a novel perspective based on feature invariance for label-specific feature learning. Specifically, invariance is injected into classifiers with respect to label-irrelevant features by intentionally manipulating these features for each class label. Accordingly, an invariance-based label-specific feature learning method, termed INVA, is proposed. INVA estimates the feature covariance matrix for each label to capture intra-class variation, thus identifying label-irrelevant features. Classifiers are then endowed with invariance to these features by solving a perturbation risk minimization problem. Furthermore, an upper bound of the perturbation risk is derived to enhance computational efficiency. Comprehensive experiments on standard multi-label benchmark datasets demonstrate the effectiveness of the proposed method.

Key words: multi-label classification; label-specific feature; feature invariance

* 基金项目: 国家自然科学基金 (62225602)
收稿时间: 2024-12-08; 修改时间: 2025-02-24; 采用时间: 2025-04-09; jos 在线出版时间: 2025-09-10
CNKI 网络首发时间: 2025-09-11

多标记分类允许每个样本同时与多个标记相关联[1,2], 能够解决真实世界中广泛存在的多义性对象的机器学习建模问题. 因而, 多标记分类在文档语义分析[3]、视觉目标识别[4]、生物信息挖掘[5]、医学疾病诊断[6]、电子商务运营[7]等众多实际应用场景中受到了广泛的关注.

多标记分类中, 一种直接的解决策略是在样本的同一特征之上导出各个标记的判别模型. 虽然简单可行, 但该策略在多义性对象建模能力上具有局限性. 这是因为多义性对象的各个标记具有潜在不同的判别偏好[8], 即各个标记与特征间的关联关系可能是不同的. 例如, 在图像识别任务中, 形状特征是“山峰”目标识别的相关特征, 而纹理特征则是“河流”目标识别的重要特征; 在文档分类任务中, “股票”“利润”等词汇特征对于“经济”主题的判别更为关键, 而“篮球”“马拉松”等词汇特征则对于“体育”主题的判别更为重要. 上述策略在同一特征上进行标记判别, 未考虑标记间存在的判别偏好差异化问题, 建模能力有限.

针对上述问题, 类属特征策略[8−11]允许不同标记的判别过程使用不同的特征, 通过特征定制充分兼顾每个标记的判别偏好, 从而有效改善多标记分类性能. 为学习类属特征, 已有方法关注于利用特征处理技术对样本中标记判别的相关特征进行提取. 其中, 一类重要的方法利用特征变换技术构造类属特征. 代表性方法 LIFT[8]首先对各个标记的正、负示例进行聚类分析, 然后将聚类原型作为特征变换基实现类属特征变换. 另一类方法则利用特征选择技术构造类属特征. 代表性方法 LLSF[9]在 LASSO 回归[12]框架下为各个标记分别选择判别过程的相关特征子集.

与上述直接提取标记相关特征的常规做法不同, 本文尝试通过对标记判别的无关特征进行操纵实现类属特征学习. 在这个研究角度上, DELA 方法[13]是目前仅有的一项工作. 在随机特征扰动框架下, 该方法根据模型输出关于输入扰动的灵敏度差异辨识各个标记的无关特征, 通过求解概率放松的扰动风险最小化问题赋予分类模型关于无关特征的不变性. 然而, 学习过程中输入-输出灵敏度差异动态变化, 无法保证无关特征辨识的可靠性, 致使该方法在复杂学习场景中表现不佳.

基于上述考虑, 本文提出了一种基于不变性注入的多标记类属特征学习方法 INVA (multi-label label-specific feature learning based on invariance injection). 在 DELA 方法基础上, INVA 方法通过显式建模数据分布特性, 增强无关特征辨识过程的可靠性. 具体而言, INVA 方法对各个标记的特征协方差矩阵进行估计, 该二阶矩统计量反映了类内各维度特征的变化情况[14], 为各个标记无关特征的辨识提供了统计依据: 其中, 类内变化程度较大的特征分量, 对应标记判别的无关特征; 而类内变化程度较小的特征分量, 则对应标记判别的相关特征. 进而, INVA 方法根据特征协方差矩阵生成随机噪声, 对标记判别的无关特征进行扰动, 从而通过求解定义在扰动训练样例之上的扰动风险最小化问题导出具有无关特征不变性的分类模型. 文中进一步推导了扰动风险最小化问题的理论上界, 该上界规避了最小化问题求解中难处理的期望运算, 提高了方法的计算效率. 多标记基准数据集上全面的实验分析结果表明, 所提方法能够有效提升多标记分类性能.

本文第 1 节简要地回顾相关的研究工作. 第 2 节介绍提出的 INVA 方法的技术细节. 第 3 节展示基准数据集上全面的实验分析结果. 第 4 节对全文进行总结.

1 相关工作

1.1 多标记分类

多标记分类问题在学术界得到了广泛的关注[1,2]. 由于多标记分类的输出空间具有关于标记数量的指数级规模, 建模标记间的依赖关系是多标记分类已有研究体系中的关键问题之一. 根据方法设计中考虑的标记相关性阶数的不同, 已有方法可分为 3 类, 分别是一阶方法[15,16]、二阶方法[17,18]和高阶方法[19−22]. 近年来, 许多研究工作致力于结合深度学习技术对已有的标记依赖关系建模方法进行改进. 例如, 一些工作尝试引入循环神经网络[23,24]、图神经网络[25,26]等深度模型来建模标记间的依赖关系. 标记嵌入方法[27,28]则通过在深度隐空间中进行特征和标记对齐, 实现标记依赖关系的隐式建模.

标记依赖关系建模方法主要关注输出空间处理. 作为上述方法的补充, 类属特征策略旨在对多标记数据的输入空间进行处理, 从而改善多标记分类性能. 根据类属特征学习过程中使用的特征处理技术的不同, 已有方法可分

为两类, 即类属特征变换方法和类属特征选择方法.

类属特征变换方法利用特征变换技术构造类属特征. 这类方法将刻画各个标记数据分布特性的聚类原型作为特征变换基, 从而将样本投影到类属特征空间中, 实现类属特征构造. 代表性方法 LIFT[8]首先通过聚类分析技术获取各个标记的聚类原型, 然后通过查询样本与这些聚类原型间的距离构造类属特征. 后续工作尝试对 LIFT 方法的技术框架进行改进. 例如, 利用聚类集成技术[29,30]或谱聚类技术[31]缓解聚类原型过程的随机性; 引入近邻信息[32−34]或全局拓扑信息[11,35−37]对基于度量的类属特征进行增广; 借助深度生成式模型实现原型和特征构造过程的判别优化[38].

类属特征选择方法利用特征选择技术构造类属特征. 代表性方法 LLSF[9]在 LASSO 回归[12]框架下为各个标记分别选择判别过程的相关特征子集, 并引入成对标记共现关系约束嵌入式特征选择过程. 后续工作从多个角度增强 LLSF 方法. 例如, 在保留的相关特征子集上施加非稀疏性约束[39,40]、引入判别导向的正则项[41−43]以及在嵌入空间中进行特征选择[44−47]等.

上述方法均关注如何利用特征处理技术对样本中标记判别的相关特征进行提取, 忽略了无关特征操纵在学习类属特征时的潜在能力. DELA 方法[13]是目前仅有的一种基于无关特征操纵的类属特征学习方法. 在随机特征扰动框架下, 该方法根据模型输出关于输入扰动的灵敏度差异辨识各个标记的无关特征, 通过求解概率放松的扰动风险最小化问题赋予分类模型关于无关特征的不变性. 与 DELA 方法相比, 本文提出的 INVA 方法使用了基于统计矩信息的无关特征辨识技术, 通过显式建模数据分布特性增强无关特征辨识过程的可靠性. 此外, INVA 方法推导了扰动风险最小化问题的理论上界, 有效降低了问题求解的计算复杂度.

1.2 特征不变性

特征不变性是计算机视觉领域一个经典的研究课题. 特征工程时期, 一些研究试图通过人工设计具有尺度不变性[48]、旋转不变性[49]的特征. 深度学习时期的工作则直接借助数据增广[50]和表示解耦[51]等技术赋予模型关于特定干扰因子的不变性. 本文的 INVA 方法借鉴了这一思想, 并将其推广到类属特征学习问题中. 与已有工作预先指定判别过程的无关特征因子 (例如, 尺度、旋转、颜色等) 不同, INVA 方法尝试通过学习来辨识各个标记判别的无关特征.

1.3 特征扰动技术

特征扰动是机器学习中一类得到广泛应用的技术手段. 例如, Dropout 技术[52−54]通过在模型计算过程中对网络特征层的神经元施加随机失活扰动, 能够增强分布式表示各维度特征的冗余性. 而本文利用特征扰动技术旨在移除分类模型对样本表示中判别无关特征的依赖性. 对抗攻击[55,56]试图寻找最脆弱的特征方向对样本进行扰动, 从而最大化模型在扰动样本上的预测误差. 与此目标不同, 本文在标记判别的无关特征方向上对样本进行扰动, 实现扰动样本上预测风险的最小化. 此外, 模型可解释性的研究中, 特征扰动被用于模型预测的事后归因[57,58]. 而本文在学习过程中进行特征扰动, 旨在提升分类模型的泛化性能.

2 基于不变性注入的多标记类属特征学习方法 INVA

本节首先给出多标记分类问题的形式化定义, 然后对无关特征操纵视角下的 DELA 方法进行简要介绍. 在此基础上, 阐述本文提出的 INVA 方法的主要思想, 并详细介绍其技术细节.

2.1 概 览

令 $X=\mathbb{R}^d$ 表示输入空间, $y=\{l_1,l_2,\ldots,l_q\}$ 表示包含 q 个类别标记的标记集合. 一个多标记样例记为 $(\mathbf{x},Y)$, 其中 $\mathbf{x}\in X$ 表示样例的 d 维特征向量 (即样本), $Y\subseteq y$ 表示样例的相关标记集合. 多标记分类旨在从给定的多标记训练集 $\mathcal{D}=\{(\mathbf{x}_i,Y_i)|1\leqslant i\leqslant m\}$ 中导出一个多标记预测模型 $h:X\to 2^y$, 实现从输入空间到输出空间 (标记集合 y 的幂集) 的映射. 对于任意未见样本 $\mathbf{u}\in X$, 该预测模型能够对其相关标记进行预测, 即 $h(\mathbf{u})\in y$.

为保证本文注释系统的简洁性, 引入一个 q 维指示向量 $\mathbf{y}=[y_1,y_2,\ldots,y_q]\in\{0,1\}^q$ 来表示样例 $(\mathbf{x},Y)$ 的相关标记集合 Y. 其中, 若标记 l_k 是样例的相关标记, 即 $l_k\in Y$, 则 $y_k=1$; 否则, $y_k=0$.

在随机特征扰动框架下, DELA 方法构造了一个扰动风险最小化问题, 同时辨识各个标记判别的无关特征, 并

赋予分类模型关于无关特征变化的不变性. 扰动风险最小化问题如下所示:

$$\min_{\phi,\Theta,S}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\mathbb{E}_{p(\varepsilon)}\left[\sum_{k=1}^{q}\mathcal{L}\left(f_k\left(e_\phi(\mathbf{x})+\mathbf{i}_{S_k}\odot\varepsilon;\theta_k\right),y_k\right)\right] \tag{1}$$

其中, $\mathcal{L}:\mathbb{R}\times\mathbb{R}\to\mathbb{R}_+$ 表示二元交叉熵函数. $e_\phi:\mathbb{R}^d\to\mathbb{R}^{d_z}$ 为一个由 ϕ 参数化的嵌入函数, 用于获取样本表示 $\mathbf{z}=e_\phi(\mathbf{x})$. $f_k(\cdot;\theta_k):\mathbb{R}^{d_z}\to\mathbb{R}$ 表示标记 l_k 由 $\theta_k=\{\mathbf{w}_k\in\mathbb{R}^{d_z},b_k\in\mathbb{R}\}$ 参数化的分类模型, $\Theta=\{\theta_1,\theta_2,\ldots,\theta_q\}$ 为各个标记的分类模型参数的集合. $S=\{S_1,S_2,\ldots,S_q\}$, 其中 S_k 表示标记 l_k 的无关特征子集. $\mathbf{i}_{S_k}\in\{0,1\}^{d_z}$ 为无关特征子集 S_k 的指示向量, 其非零分量对应标记判别的无关特征索引. $\varepsilon\in\mathbb{R}^{d_z}$ 为一个随机噪声变量, 其服从一个各向同性高斯分布, 即 $p(\varepsilon)=\mathcal{N}(\varepsilon;\mathbf{0},\sigma^2\cdot\mathbf{I})$.

在公式 (1) 中, DELA 方法根据分类模型输出关于输入扰动的灵敏度差异辨识各个标记的无关特征; 同时, 通过在无关特征上注入随机噪声扰动模拟无关特征的变化, 从而在学习过程中逐渐赋予分类模型关于无关特征变化的不变性. 其中, 有效辨识标记判别的无关特征是扰动风险最小化问题求解的关键. 在复杂的多标记分类场景中, 单纯依靠输入-输出灵敏度差异区分相关特征和无关特征并不可靠. 后续章节将详细介绍本文在 DELA 方法基础上提出的改进方法.

2.2 协方差诱导的无关特征辨识

针对 DELA 方法的不足, 本文提出了一种特征协方差诱导的无关特征辨识技术, 通过显式建模数据分布特性, 增强无关特征辨识过程的可靠性:

$$\min_{\phi,\Theta}\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\mathbb{E}_{p(\varepsilon_k)}\left[\mathcal{L}\left(f_k\left(e_\phi(\mathbf{x})+\varepsilon_k;\theta_k\right),y_k\right)\right] \tag{2}$$

其中, 随机噪声变量 $\varepsilon_k\in\mathbb{R}^{d_z}$ 服从一个零均值高斯分布, 即 $p(\varepsilon_k)=\mathcal{N}(\varepsilon_k;\mathbf{0},\lambda\cdot\boldsymbol{\Sigma}_k)$. $\boldsymbol{\Sigma}_k\in\mathbb{R}_+^{d_z\times d_z}$ 表示标记 l_k 的特征协方差矩阵, 从标记 l_k 的正例中统计得到, $\lambda>0$ 为控制特征扰动规模的强度因子. 该二阶矩统计量反映了类内各维度特征的变化情况, 为各个标记无关特征的辨识提供了统计依据: 其中, 类内变化程度较大的特征分量, 对应标记判别的无关特征; 而类内变化程度较小的特征分量, 则对应标记判别的相关特征. 同时, 特征协方差矩阵刻画了各维度特征间的关联关系, 因而噪声扰动 ε_k 能够更好地模拟真实数据中的特征变化情况. 例如, 人的“年龄”大小与“皱纹”深浅之间存在着一定的关联关系, 随着“年龄”增长, “皱纹”通常会加深. 特征协方差矩阵能够避免噪声扰动 ε_k 单一地改变“年龄”大小, 而不改变“皱纹”深浅.

由于嵌入函数 e_ϕ 在学习过程中不断优化, 训练样本的表示 $\mathbf{z}=e_\phi(\mathbf{x})$ 也随之变化. 因此, 在公式 (2) 所示的扰动风险最小化问题的求解过程中, 需要不断更新各个标记的特征协方差矩阵. 为了实现的高效性, 采用了在线估计的方式对特征协方差矩阵进行更新. 具体而言, 对于第 t 次优化迭代的训练样本子集 $\mathcal{B}$, 统计样本子集的特征均值和特征协方差矩阵:

$$\begin{cases}\tilde{\mu}_k^{(t)}=\dfrac{1}{|\mathcal{B}_k|}\displaystyle\sum_{\mathbf{x}\in\mathcal{B}_k}e_\phi(\mathbf{x})\\ \tilde{\boldsymbol{\Sigma}}_k^{(t)}=\dfrac{1}{|\mathcal{B}_k|}\displaystyle\sum_{\mathbf{x}\in\mathcal{B}_k}\left(e_\phi(\mathbf{x})-\tilde{\mu}_k^{(t)}\right)\left(e_\phi(\mathbf{x})-\tilde{\mu}_k^{(t)}\right)^{\mathrm{T}}\end{cases} \tag{3}$$

其中, $\mathcal{B}_k$ 为训练样本子集 $\mathcal{B}$ 中标记 l_k 的正例集合. 记第 t 次优化迭代中标记 l_k 的正例数量为 $m_k^{(t)}$ (有 $m_k^{(t)}=|\mathcal{B}_k|$), 前 $(t-1)$ 次优化迭代中出现的标记 l_k 的正例总数为 $n_k^{(t-1)}$, 可按如下方式对各个标记的特征均值和特征协方差矩阵进行更新:

$$\begin{cases}\mu_k^{(t)}=\dfrac{n_k^{(t-1)}\mu_k^{(t-1)}+m_k^{(t)}\tilde{\mu}_k^{(t)}}{n_k^{(t-1)}+m_k^{(t)}}\\ \boldsymbol{\Sigma}_k^{(t)}=\dfrac{n_k^{(t-1)}\boldsymbol{\Sigma}_k^{(t-1)}+m_k^{(t)}\tilde{\boldsymbol{\Sigma}}_k^{(t)}}{n_k^{(t-1)}+m_k^{(t)}}+\dfrac{n_k^{(t-1)}m_k^{(t)}\left(\mu_k^{(t-1)}-\tilde{\mu}_k^{(t)}\right)\left(\mu_k^{(t-1)}-\tilde{\mu}_k^{(t)}\right)^{\mathrm{T}}}{\left(n_k^{(t-1)}+m_k^{(t)}\right)^2}\\ n_k^{(t)}=n_k^{(t-1)}+m_k^{(t)}\end{cases} \tag{4}$$

2.3 扰动风险最小化问题上界推导

在公式 (2) 所示的扰动风险最小化问题中, 涉及关于随机噪声变量的期望计算. 该期望项难以进行解析计算, 常规做法是利用蒙特卡洛采样技术对其进行估计:

$$\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\mathbb{E}_{p(\varepsilon_k)}[\mathcal{L}(f_k(e_\phi(\mathbf{x})+\varepsilon_k;\theta_k),y_k)]\approx\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\frac{1}{L}\sum_{i=1}^{L}\left[\mathcal{L}\left(f_k\left(e_\phi(\mathbf{x})+\varepsilon_k^{(i)};\theta_k\right),y_k\right)\right] \tag{5}$$

其中, 需要对随机噪声变量 ε_k 进行 L 次采样, 然后分别对样本表示进行扰动, 并由分类模型进行预测. 上述常规实现具有关于采样次数 L 的线性复杂度, 当采样次数 L 趋向于无穷大时, 虽能对原期望项进行精确估计, 但计算开销过大.

为提高方法的计算效率, 本节进一步推导了原期望项的理论上界:

$$\begin{aligned}&\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\mathbb{E}_{p(\varepsilon_k)}[\mathcal{L}(f_k(e_\phi(\mathbf{x})+\varepsilon_k;\theta_k),y_k)]\\&\leqslant\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\left[-y_k\log\rho\left(f_k(e_\phi(\mathbf{x});\theta_k)-\frac{\lambda}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}\right)-(1-y_k)\log\left(1-\rho\left(f_k(e_\phi(\mathbf{x});\theta_k)+\frac{\lambda}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}\right)\right)\right]\end{aligned} \tag{6}$$

其中, $\rho(\cdot)$ 表示 Sigmoid 函数. 与原期望项相比, 上述上界在计算过程中, 仅需在分类模型输出的逻辑值上附加由特征协方差矩阵导出的修正值, 无需再对随机噪声变量 ε_k 进行多次采样、计算, 从而有效提高了方法的计算效率.

公式 (6) 的推导过程如下:

$$\begin{aligned}&\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\mathbb{E}_{p(\varepsilon_k)}[\mathcal{L}(f_k(e_\phi(\mathbf{x})+\varepsilon_k;\theta_k),y_k)]\\&=\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\mathbb{E}_{p(\varepsilon_k)}[-y_k\log\rho(f_k(\mathbf{z}_k;\theta_k))-(1-y_k)\log(1-\rho(f_k(\mathbf{z}_k;\theta_k)))]\\&\leqslant\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\left[y_k\log\mathbb{E}_{p(\varepsilon_k)}\left(\frac{1}{\rho(f_k(\mathbf{z}_k;\theta_k))}\right)+(1-y_k)\log\mathbb{E}_{p(\varepsilon_k)}\left(\frac{1}{1-\rho(f_k(\mathbf{z}_k;\theta_k))}\right)\right]\\&=\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\left[y_k\log\left(1+\mathbb{E}_{p(\varepsilon_k)}\left(\mathrm{e}^{-\mathbf{w}_k^{\mathrm{T}}\mathbf{z}_k-b}\right)\right)+(1-y_k)\log\left(1+\mathbb{E}_{p(\varepsilon_k)}\left(\mathrm{e}^{\mathbf{w}_k^{\mathrm{T}}\mathbf{z}_k+b}\right)\right)\right]\\&=\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\left[y_k\log\left(1+\mathrm{e}^{-\mathbf{w}_k^{\mathrm{T}}e_\phi(\mathbf{x})-b+\frac{1}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}}\right)+(1-y_k)\log\left(1+\mathrm{e}^{\mathbf{w}_k^{\mathrm{T}}e_\phi(\mathbf{x})+b+\frac{1}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}}\right)\right]\\&=\mathbb{E}_{p(\mathbf{x},\mathbf{y})}\sum_{k=1}^{q}\left[-y_k\log\rho\left(f_k(e_\phi(\mathbf{x});\theta_k)-\frac{\lambda}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}\right)-(1-y_k)\log\left(1-\rho\left(f_k(e_\phi(\mathbf{x});\theta_k)+\frac{\lambda}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}\right)\right)\right]\end{aligned} \tag{7}$$

其中, 记 $\mathbf{z}_k=e_\phi(\mathbf{x})+\varepsilon_k$. 推导过程中, 不等式关系由 Jensen 不等式导出, 即 $\mathbb{E}[\log X]\leqslant\log\mathbb{E}[X]$. 倒数第 2 个等式的推导使用了高斯分布的特性, 即:

$$\mathbb{E}_{p(X)}\left[\mathrm{e}^{tX}\right]=\mathrm{e}^{t\mu+\frac{1}{2}\sigma^2t^2},\ X\sim\mathcal{N}\left(X;\mu,\sigma^2\right) \tag{8}$$

由于 $\kappa=\mathbf{w}_k^{\mathrm{T}}\mathbf{z}_k+b\sim\mathcal{N}(\kappa;\mathbf{w}_k^{\mathrm{T}}e_\phi(\mathbf{x})+b,\lambda\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w})$ 为一高斯随机变量, 因此有:

$$\begin{cases}\mathbb{E}_{p(\varepsilon_k)}\left(\mathrm{e}^{\mathbf{w}_k^{\mathrm{T}}\mathbf{z}_k+b}\right)=\mathrm{e}^{\mathbf{w}_k^{\mathrm{T}}e_\phi(\mathbf{x})+b+\frac{1}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}}\\\mathbb{E}_{p(\varepsilon_k)}\left(\mathrm{e}^{-\mathbf{w}_k^{\mathrm{T}}\mathbf{z}_k-b}\right)=\mathrm{e}^{-\mathbf{w}_k^{\mathrm{T}}e_\phi(\mathbf{x})-b+\frac{1}{2}\mathbf{w}_k^{\mathrm{T}}\boldsymbol{\Sigma}_k\mathbf{w}}\end{cases} \tag{9}$$

基于导出的理论上界, INVA 方法中扰动风险最小化问题的求解过程如伪代码 1 所示.

伪代码 1. INVA 方法优化过程.

输入: 多标记训练集 $\mathcal{D}$, 强度因子 λ;

1. 随机初始化模型参数 ϕ 和 Θ;

2. **for** t=0 **to** T **do**

3. 从 $\mathcal{D}$ 中随机采样一个子集 $\mathcal{B}$;
4. 更新特征协方差矩阵 $\boldsymbol{\Sigma}_k$ $(1 \leqslant k \leqslant q)$ (公式 (4));
5. 计算扰动风险最小化问题上界 (公式 (6));
6. 利用梯度下降法更新模型参数 ϕ 和 Θ;
7. **end for**

输出: 模型参数 ϕ 和 Θ.

3 实 验

3.1 实验设置

数据集: 实验中共使用了 8 个多标记基准数据集, 以进行全面的性能评价. 表 1 展示了各个数据集所具有的多样化的多标记特性, 包括样例数量、特征维度、标记数量、特征类型、标记势 (即每个样本具有的相关标记数量均值). 参照文献 [8], 对数据集 rcv-s1 和 tmc2007 进行降维, 根据字典频次保留前 2% 的特征. 数据集 mirflickr 使用局部描述子 DenseSift 提取的特征.

表 1 多标记基准数据集特性表

数据集	样例数量	特征维度	标记数量	特征类型	标记势	领域
yeast	2417	103	14	数值型	4.237	生物[1]
rcv1-s1	6000	944	101	数值型	2.880	文本[1]
Corel16k-s1	13766	500	153	类别型	2.859	图像[1]
delicious	16105	500	983	类别型	19.020	文本[1]
mirflickr	25000	1000	38	数值型	4.716	图像[2]
tmc2007	28596	981	22	类别型	2.158	文本[1]
mediamill	43907	120	101	数值型	4.376	视频[1]
bookmarks	87856	2150	208	类别型	2.028	文本[1]

注: 领域列中, 1的数据集链接为http://mulan.sourceforge.net/datasets.html, 2的数据集链接为http://lear.inrialpes.fr/people/guillaumin/data.php

评价指标: 为全面评价算法的多标记分类性能, 实验中使用了 6 项常用的多标记评价指标, 包括 Average precision、Macro-averaging AUC、Hamming loss、One-error、Coverage 和 Ranking loss. 评价指标的定义参见文献 [1].

实现细节: 本文采用与 DELA 方法[13]相同的模型结构和优化方法. 具体而言, 嵌入函数 e_ϕ 被实现为全连接神经网络, 隐层维度设为 [256, 512, 256]. 使用 Adam 优化器进行网络参数优化, 批样本大小设为 128、权重衰减因子设为 1E–4、动量因子设为 0.999 和 0.9.

3.2 对比分析

INVA 方法与 6 个代表性的多标记分类算法进行了性能比较. 实验中, 对比算法使用原始文献中提供的推荐超参数配置.

- LIFT[8]: 一个基于原型投影的类属特征变换方法, [$r = 0.1$].
- LLSF[9]: 一个 LASSO 框架下的类属特征选择方法, 考虑成对标记间的共现关系, [网格超参数搜索: $\alpha,\beta \in \{2^{-10},2^{-9},\ldots,2^{10}\}$, $\gamma = 0.01$].
- C2AE[27]: 一个深度标记嵌入方法, 利用深度典型相关分析 (canonical correlation analysis) 技术和自编码器将特征和标记嵌入到同一空间中, [超参数搜索: $\alpha \in \{0.1, 1, 2, 5, 10\}$].
- MPVAE[28]: 在概率隐空间中, 利用变分自编码器对齐样例的特征和标记. 同时, 学习协方差矩阵建模标记依赖关系, [$\lambda_1 = \lambda_2 = 0.5$, $\lambda_3 = 10$, $\beta = 1.1$].

● CLIF[45]: 一个深度类属特征学习方法, 在合作学习框架下对标记依赖关系和类属特征进行协同优化, [网格超参数搜索: $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 1, 2, 5, 10\}$, $d_e \in \{64, 128, 256\}$].

● DELA[13]: 无关特征操纵视角下的深度类属特征学习方法, [超参数搜索: $\beta \in \{10^{-5}, 10^{-4}, \ldots, 10\}$].

INVA 方法包含一个超参数, 即公式 (2) 中的强度因子 λ. 实验中对其进行搜索, 搜索范围设为{0.01, 0.1, 0.25, 0.5, 0.75, 1.0, 2.5}. 为性能评价的公平起见, 所有的深度方法均使用相同的神经网络结构, 并搜索最佳的学习率和学习率衰减机制. 每个数据集中, 随机抽取 10% 的样例作为验证集, 用于超参数搜索. 在剩余的 90% 样例上, 进行十折交叉验证来评价算法性能.

表 2 和表 3 汇报了关于各项评价指标的详细实验结果. 对于各项评价指标, 表中使用符号↑(↓) 表示评价指标取值越大 (越小), 性能越好. 对比分析中取得的最佳性能在表中加粗显示. 此外, 本节进行显著度为 0.05 的威尔科克森符号秩检验 (Wilcoxon signed-ranks test)[59], 来分析 INVA 方法相较于已有方法是否展现出统计上更优的性能. 表 4 汇总了各项评价指标上的统计检验结果.

表 2　INVA 方法和对比方法的 Average precision、Macro-averaging AUC 和 Hamming loss 性能

指标	数据集	LIFT	LLSF	C2AE	MPVAE	CLIF	DELA	INVA
Average precision ↑	yeast	0.7680±0.0253	0.7564±0.0229	0.7389±0.0217	0.7641±0.0252	0.7650±0.0215	0.7691±0.0227	**0.7757±0.0229**
	rcv1-s1	0.5921±0.0145	0.6129±0.0116	0.6147±0.0100	0.6332±0.0136	0.6246±0.0094	0.6391±0.0153	**0.6445±0.0136**
	Corel16k-s1	0.3168±0.0059	0.3428±0.0053	0.3297±0.0042	0.3646±0.0063	0.3516±0.0070	0.3675±0.0062	**0.3724±0.0045**
	delicious	0.3833±0.0061	0.3587±0.0079	0.3648±0.0075	0.4042±0.0065	0.3832±0.0069	0.4082±0.0053	**0.4092±0.0062**
	mirflickr	0.6516±0.0041	0.6477±0.0039	0.6627±0.0064	0.6849±0.0053	0.6857±0.0025	0.6960±0.0043	**0.6973±0.0027**
	tmc2007	0.8207±0.0048	0.8130±0.0050	0.7977±0.0048	0.8297±0.0032	0.8189±0.0024	0.8363±0.0037	**0.8371±0.0038**
	mediamill	0.7417±0.0058	0.7275±0.0051	0.7266±0.0051	0.7669±0.0062	0.7650±0.0061	0.7883±0.0049	**0.7965±0.0046**
	bookmarks	0.5119±0.0044	0.4920±0.0035	0.4707±0.0046	0.5104±0.0050	0.4928±0.0036	0.5191±0.0036	**0.5203±0.0034**
Macro-averaging AUC ↑	yeast	0.6716±0.0173	0.6640±0.0209	0.6931±0.0183	0.7091±0.0209	0.7115±0.0185	**0.7335±0.0168**	0.7293±0.0164
	rcv1-s1	0.9241±0.0103	0.9062±0.0100	0.9131±0.0084	0.9368±0.0078	0.9320±0.0044	0.9374±0.0079	**0.9382±0.0074**
	Corel16k-s1	0.6937±0.0097	0.6614±0.0075	0.7212±0.0131	0.7867±0.0130	0.7657±0.0105	0.7872±0.0098	**0.7883±0.0106**
	delicious	0.7919±0.0044	0.7509±0.0047	0.7830±0.0052	0.8272±0.0039	0.8107±0.0043	**0.8305±0.0030**	0.8284±0.0037
	mirflickr	0.8091±0.0077	0.8196±0.0042	0.8213±0.0046	0.8461±0.0040	0.8436±0.0045	0.8538±0.0045	**0.8550±0.0050**
	tmc2007	0.9229±0.0035	0.9225±0.0040	0.8993±0.0052	0.9307±0.0038	0.9274±0.0048	0.9356±0.0037	**0.9360±0.0042**
	mediamill	0.8302±0.0080	0.7874±0.0110	0.8172±0.0069	0.8627±0.0083	0.8703±0.0086	0.8836±0.0067	**0.8883±0.0066**
	bookmarks	0.8984±0.0030	0.8857±0.0037	0.8403±0.0040	0.9106±0.0015	0.9024±0.0028	0.9117±0.0022	**0.9149±0.0032**
Hamming loss ↓	yeast	0.1927±0.0112	0.2019±0.0110	0.2212±0.0142	0.2140±0.0098	0.1963±0.0114	0.2013±0.0123	**0.1917±0.0099**
	rcv1-s1	**0.0259±0.0009**	0.0263±0.0011	0.0408±0.0016	0.0270±0.0011	0.0267±0.0011	0.0266±0.0009	0.0264±0.0012
	Corel16k-s1	0.0187±0.0002	0.0186±0.0002	0.0233±0.0005	0.0188±0.0003	0.0188±0.0003	0.0186±0.0002	**0.0185±0.0002**
	delicious	0.0180±0.0001	0.0184±0.0002	0.0248±0.0006	**0.0177±0.0001**	0.0179±0.0001	0.0178±0.0001	0.0178±0.0001
	mirflickr	0.1019±0.0009	0.1005±0.0009	0.1259±0.0037	0.0969±0.0011	0.0965±0.0008	0.0945±0.0012	**0.0943±0.0009**
	tmc2007	0.0603±0.0007	0.0607±0.0013	0.0632±0.0014	0.0586±0.0006	0.0587±0.0010	0.0572±0.0009	**0.0571±0.0009**
	mediamill	0.0291±0.0003	0.0304±0.0002	0.0348±0.0004	0.0281±0.0004	0.0279±0.0004	0.0260±0.0004	**0.0252±0.0004**
	bookmarks	0.0086±0.0001	0.0087±0.0001	0.0106±0.0001	0.0087±0.0001	**0.0085±0.0001**	0.0086±0.0001	0.0086±0.0001

基于上述实验结果, 可以发现:

● 如表 2 和表 3 所示, 在共计 48 种实验配置 (8 个数据集×6 项评价指标) 中, INVA 方法取得最佳性能共计 39 次 (81%), 清晰展示了其优异的多标记分类性能.

● 表 4 显示, INVA 方法在各项评价指标上显著优于深度标记嵌入方法 C2AE 和 MPVAE. C2AE 和 MPVAE 关注通过标记嵌入技术建模标记依赖关系. INVA 方法优于 C2AE 和 MPVAE 方法的性能, 充分证明了类属特征是一种改善多标记分类性能的有效策略.

● 同时, 相较于其他的类属特征学习方法, INVA 方法取得了显著更佳的性能, 这证明了从无关特征操纵角度进行类属特征学习的有效性. 特别地, 与同样基于无关特征操纵的 DELA 方法相比, INVA 方法在各个数据集上展现出一致更优的性能, 证明了 INVA 方法在技术上的优越性.

表 3 INVA 方法和对比方法的 One-error、Coverage 和 Ranking loss 性能

指标	数据集	LIFT	LLSF	C2AE	MPVAE	CLIF	DELA	INVA
One-error ↓	yeast	**0.2193±0.0408**	0.2294±0.0352	0.2689±0.0309	0.2211±0.0389	0.2321±0.0323	0.2349±0.0309	0.2261±0.0382
	rcv1-s1	0.4106±0.0194	0.4220±0.0161	0.4383±0.0210	0.4078±0.0330	0.4102±0.0192	0.4006±0.0215	**0.3885±0.0192**
	Corel16k-s1	0.6764±0.0126	0.6398±0.0092	0.6442±0.0090	0.6331±0.0155	0.6420±0.0126	0.6268±0.0099	**0.6193±0.0133**
	delicious	0.3339±0.0153	0.3537±0.0145	0.3374±0.0178	0.3070±0.0183	0.3194±0.0180	**0.3061±0.0138**	0.3072±0.0109
	mirflickr	0.3076±0.0106	0.3025±0.0089	0.2848±0.0110	0.2740±0.0105	0.2702±0.0083	0.2622±0.0089	**0.2600±0.0119**
	tmc2007	0.2125±0.0076	0.2245±0.0094	0.2296±0.0082	0.2031±0.0072	0.2003±0.0036	0.1945±0.0067	**0.1942±0.0081**
	mediamill	0.1757±0.0122	0.1590±0.0040	0.1643±0.0072	0.1422±0.0046	0.1421±0.0060	0.1288±0.0038	**0.1252±0.0041**
	bookmarks	0.5115±0.0044	0.5319±0.0054	0.5408±0.0070	0.5165±0.0068	0.5337±0.0050	0.5079±0.0042	**0.5053±0.0046**
Coverage ↓	yeast	0.4518±0.0190	0.4626±0.0193	0.4740±0.0166	0.4503±0.0201	0.4511±0.0159	0.4388±0.0179	**0.4357±0.0175**
	rcv1-s1	0.1231±0.0124	0.1245±0.0120	0.1040±0.0078	0.0932±0.0091	0.0992±0.0068	0.0866±0.0085	**0.0818±0.0071**
	Corel16k-s1	0.3247±0.0050	0.3243±0.0071	0.3049±0.0068	0.2372±0.0055	0.2499±0.0067	0.2330±0.0049	**0.2320±0.0047**
	delicious	0.4809±0.0132	0.6150±0.0093	0.5108±0.0062	0.4058±0.0063	0.4208±0.0051	0.3943±0.0049	**0.3941±0.0062**
	mirflickr	0.3086±0.0038	0.3205±0.0043	0.3075±0.0041	0.2741±0.0031	0.2757±0.0033	0.2686±0.0046	**0.2683±0.0032**
	tmc2007	0.1193±0.0028	0.1270±0.0025	0.1511±0.0059	0.1144±0.0023	0.1161±0.0026	0.1110±0.0023	**0.1107±0.0023**
	mediamill	0.1517±0.0088	0.1671±0.0031	0.1760±0.0029	0.1233±0.0033	0.1239±0.0030	0.1143±0.0028	**0.1121±0.0028**
	bookmarks	0.1293±0.0060	0.1510±0.0032	0.1905±0.0040	0.1189±0.0028	0.1270±0.0019	0.1105±0.0019	**0.1103±0.0027**
Ranking loss ↓	yeast	0.1645±0.0175	0.1746±0.0175	0.1894±0.0151	0.1665±0.0183	0.1662±0.0140	0.1652±0.0165	**0.1592±0.0170**
	rcv1-s1	0.0490±0.0056	0.0497±0.0046	0.0428±0.0032	0.0390±0.0041	0.0426±0.0025	0.0344±0.0035	**0.0324±0.0029**
	Corel16k-s1	0.1636±0.0017	0.1611±0.0042	0.1638±0.0052	0.1239±0.0038	0.1303±0.0043	0.1202±0.0028	**0.1201±0.0029**
	delicious	0.0985±0.0020	0.1449±0.0046	0.1197±0.0021	0.0884±0.0019	0.0933±0.0017	0.0856±0.0016	**0.0847±0.0019**
	mirflickr	0.1125±0.0020	0.1196±0.0028	0.1120±0.0038	0.0939±0.0015	0.0986±0.0015	0.0937±0.0019	**0.0927±0.0013**
	tmc2007	0.0449±0.0019	0.0489±0.0017	0.0629±0.0023	0.0416±0.0013	0.0432±0.0014	**0.0395±0.0015**	0.0396±0.0017
	mediamill	0.0412±0.0017	0.0496±0.0011	0.0537±0.0014	0.0342±0.0012	0.0343±0.0011	0.0309±0.0009	**0.0303±0.0011**
	bookmarks	0.0813±0.0037	0.0947±0.0025	0.1271±0.0028	0.0767±0.0020	0.0838±0.0013	**0.0700±0.0014**	0.0711±0.0018

表 4 INVA 方法和对比方法之间的威尔科克森符号秩检验结果 (显著度为 0.05)

指标	LIFT	LLSF	C2AE	MPVAE	CLIF	DELA
Average precision	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]
Macro-averaging AUC	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	tie [0.460 9]
Hamming loss	tie [0.078 1]	win [0.031 3]	win [0.007 8]	win [0.023 4]	win [0.023 4]	win [0.031 3]
One-error	win [0.023 4]	win [0.007 8]	win [0.007 8]	win [0.039 1]	win [0.007 8]	win [0.023 4]
Coverage	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]
Ranking loss	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	win [0.007 8]	tie [0.156 3]

3.3 消融分析

本节对 INVA 方法设计中的关键元素进行消融分析. 在第 3.1.1 节中介绍的 8 个基准数据集上, 进行十折交叉验证来评价 INVA 方法和其变种方法的性能 (实验结果见表 5), 并进行威尔科克森符号秩检验以分析方法的统计性能差异 (分析结果见表 6).

无关特征操纵有效性: 在学习过程中, INVA 方法对各个标记判别的无关特征进行辨识, 并对无关特征进行扰动, 从而面向各个标记的判别偏好实现特征不变性的注入. 为验证无关特征操纵过程的有效性, 实现了两个变种方法 (分别记为 INVA-sn 和 INVA-nn). INVA-sn 方法不再辨识类属的无关特征, 仅利用标记间共享的随机噪声对样本表示 $\mathbf{z}$ 进行扰动. INVA-nn 方法进一步移除无关特征操纵过程, 直接基于标记间共享的样本表示 $\mathbf{z}$ 进行标记的判别. 如表 5 和表 6 所示, 在各项评价指标上 INVA 方法均显著优于变种方法 INVA-sn 和 INVA-nn, 证明了 INVA 方法中基于不变性注入的无关特征操纵过程在多标记分类问题中的优越性.

问题上界优化有效性: INVA 方法推导了扰动风险最小化问题的理论上界, 通过优化该上界实现扰动风险最小化问题的求解. 为验证该优化策略的有效性, 实现了一个变种方法 (记为 INVA-sa). 该方法利用公式 (5) 所示的蒙特卡洛采样技术对扰动风险最小化问题中的期望项进行估计, 通过对期望估计进行优化实现扰动风险最小化问

题的求解. 实验中, 将采样次数 L 分别设为{1, 2, 5}. 如表 5 和表 6 所示, 当 $L=1$ 时, 虽然上界优化与估计优化具有相近的计算复杂度, 但上界优化展现出更佳的泛化性能. 当 $L>1$, 估计优化的泛化性能相较于 $L=1$ 时略有提升, 但在统计上仍然差于 INVA 方法使用的上界优化技术.

表 5 INVA 方法和其变种方法的 Average precision 性能 (↑)

数据集	LIFT	INVA-sn	INVA-nn	INVA-sa ($L=1$)	INVA-sa ($L=2$)	INVA-sa ($L=5$)
yeast	**0.7757±0.0229**	0.7701±0.0214	0.7704±0.0209	0.7703±0.0261	0.7708±0.0229	0.7716±0.0224
rcv1-s1	**0.6445±0.0136**	0.6350±0.0143	0.6323±0.0134	0.6420±0.0136	0.6420±0.0127	0.6422±0.0141
Corel16k-s1	**0.3724±0.0045**	0.3709±0.0055	0.3581±0.0080	0.3682±0.0054	0.3701±0.0055	0.3719±0.0060
delicious	**0.4092±0.0062**	0.4080±0.0062	0.4076±0.0067	0.4075±0.0059	0.4076±0.0061	0.4084±0.0063
mirflickr	**0.6973±0.0027**	0.6964±0.0040	0.6966±0.0033	0.6964±0.0034	0.6960±0.0029	0.6968±0.0032
tmc2007	0.8371±0.0038	0.8359±0.0033	0.8354±0.0040	0.8357±0.0037	0.8365±0.0042	**0.8374±0.0044**
mediamill	0.7965±0.0046	0.7971±0.0050	0.7959±0.0045	0.7959±0.0049	0.7963±0.0043	**0.7966±0.0039**
bookmarks	**0.5203±0.0034**	0.5108±0.0039	0.5186±0.0038	0.5166±0.0036	0.5178±0.0037	0.5195±0.0037

注: ↑表示评价指标取值越大, 性能越好, 最佳性能加粗显示

表 6 INVA 方法和其变种方法之间的威尔科克森符号秩检验结果 (显著度为 0.05)

指标	INVA-sn	INVA-nn	INVA-sa ($L=1$)	INVA-sa ($L=2$)	INVA-sa ($L=5$)
Average precision	win [0.0156]	win [0.0078]	win [0.0078]	win [0.0078]	win [0.0313]
Macro-averaging AUC	win [0.0078]	win [0.0078]	win [0.0313]	win [0.0469]	win [0.0391]
Hamming loss	tie [0.0625]	win [0.0078]	win [0.0313]	win [0.0313]	win [0.0313]
One-error	win [0.0391]	win [0.0156]	win [0.0391]	win [0.0234]	win [0.0156]
Coverage	win [0.0469]	win [0.0234]	win [0.0469]	win [0.0391]	win [0.0156]
Ranking loss	win [0.0469]	win [0.0313]	win [0.0313]	win [0.0469]	win [0.0078]

3.4 参数敏感性分析

本节对 INVA 方法的参数敏感性进行分析. 分析实验中, 改变强度因子 λ 的取值, 观察导出模型泛化性能的变化. 相应地, 图 1 展示了 INVA 方法在 λ 的不同取值下, 模型泛化性能的变化情况. 可以发现, λ 作为控制噪声扰动强度的因子, 的确影响着 INVA 方法的泛化性能. 特别地, 若完全移除噪声扰动过程 (即当 $\lambda=0$ 时), INVA 方法的泛化性能将出现显著退化.

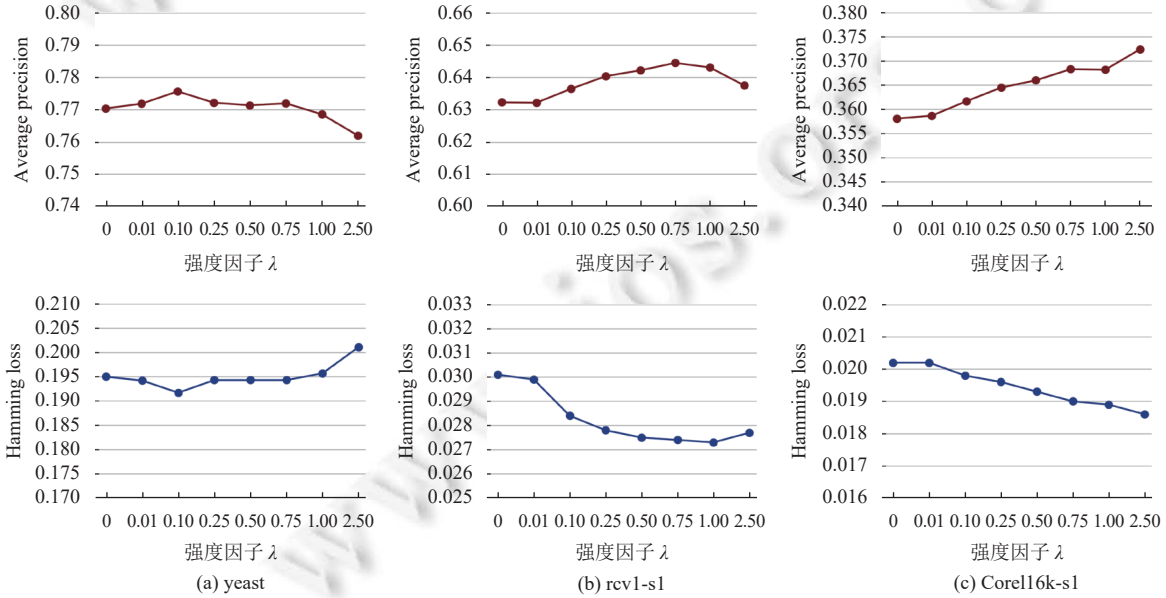

图 1 INVA 方法的参数敏感性曲线

3.5 时间复杂度分析

令 b 表示学习过程中每次优化迭代的训练样本子集 $\mathcal{B}$ 的大小, $\hat{d}$ 表示网络隐层维度的代理, INVA 方法的时间复杂度为 $\mathrm{O}\left(bq\hat{d}^2\right)$. 图 2 展示了第 3.2 节中各个对比方法在训练和测试阶段的实际运行时间. 在时间开销方面, INVA 方法与已有方法可比.

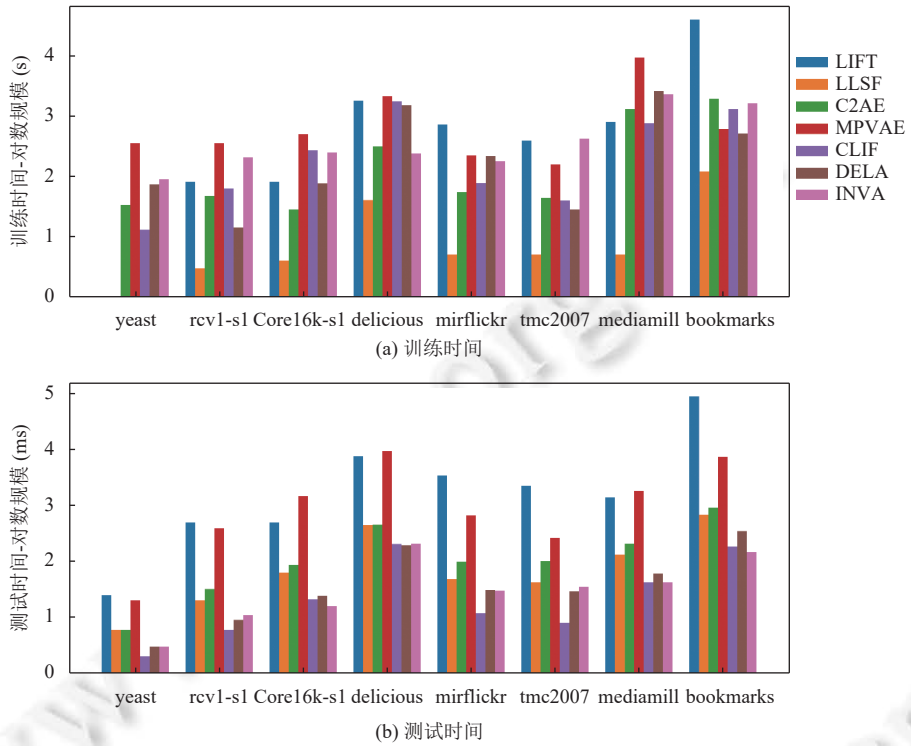

图 2 INVA 方法和对比方法的训练/测试阶段运行时间比较

4 总 结

本文提出了一种基于不变性注入的多标记类属特征学习方法 INVA. 该方法通过操纵标记判别的无关特征, 为分类模型注入关于无关特征的不变性, 从而充分地兼顾各个标记潜在不同的判别偏好. 实现中, INVA 方法通过估计特征协方差矩阵捕获各个标记的类内特征变化, 从而辨识标记判别的无关特征; 通过构造扰动风险最小化问题并推导问题上界, 从而高效地赋予分类模型关于无关特征变化的不变性. 在特性多样化的多标记基准数据集上, 与多种已有的多标记分类方法进行了全面的对比分析, 验证了本文所提的 INVA 方法在解决多标记分类问题上的有效性.

References

[1] Zhang ML, Zhou ZH. A review on multi-label learning algorithms. IEEE Trans. on Knowledge and Data Engineering, 2014, 26(8): 1819–1837. [doi: 10.1109/TKDE.2013.39]

[2] Liu WW, Shen XB, Wang HB, Tsang IW. The emerging trends of multi-label learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7955–7974. [doi: 10.1109/TPAMI.2021.3119334]

[3] Zong DM, Sun SL. BGNN-XML: Bilateral graph neural networks for extreme multi-label text classification. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(7): 6698–6709. [doi: 10.1109/TKDE.2022.3193657]

[4] Kerrigan A, Duarte K, Rawat YS, Shah M. Reformulating zero-shot action recognition for multi-label actions. 2021. https://papers.nips.cc/

paper/2021/file/d6539d3b57159babf6a72e106beb45bd-Paper.pdf

[5] Li YX, Ji SW, Kumar S, Ye JP, Zhou ZH. Drosophila gene expression pattern annotation through multi-instance multi-label learning. IEEE/ACM Trans. on Computational Biology and Bioinformatics, 2012, 9(1): 98–112. [doi: 10.1109/TCBB.2011.73]

[6] Guan QJ, Huang YP. Multi-label chest X-ray image classification via category-wise residual attention learning. Pattern Recognition Letters, 2020, 130: 259–266. [doi: 10.1016/j.patrec.2018.10.027]

[7] Chen C, Zhang M, Zhang YF, Ma WZ, Liu YQ, Ma SP. Efficient heterogeneous collaborative filtering without negative sampling for recommendation. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 19–26. [doi: 10.1609/aaai.v34i01.5329]

[8] Zhang ML, Wu L. LIFT: Multi-label learning with label-specific features. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107–120. [doi: 10.1109/TPAMI.2014.2339815]

[9] Huang J, Li GR, Huang QM, Wu XD. Learning label-specific features and class-dependent labels for multi-label classification. IEEE Trans. on Knowledge and Data Engineering, 2016, 28(12): 3309–3323. [doi: 10.1109/TKDE.2016.2608339]

[10] Jia XY, Zhu SS, Li WW. Joint label-specific features and correlation information for multi-label learning. Journal of Computer Science and Technology, 2020, 35(2): 247–258. [doi: 10.1007/s11390-020-9900-z]

[11] Guo YM, Chung FL, Li GZ, Wang JC, Gee JC. Leveraging label-specific discriminant mapping features for multi-label learning. ACM Trans. on Knowledge Discovery from Data, 2019, 13(2): 24. [doi: 10.1145/3319911]

[12] Tibshirani R. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996, 58(1): 267–288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]

[13] Hang JY, Zhang ML. Dual perspective of label-specific feature learning for multi-label classification. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 8375–8386.

[14] Wang YL, Pan XR, Song SJ, Zhang H, Wu C, Huang G. Implicit semantic data augmentation for deep networks. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 12635–12644.

[15] Boutell MR, Luo JB, Shen XP, Brown CM. Learning multi-label scene classification. Pattern Recognition, 2004, 37(9): 1757–1771. [doi: 10.1016/j.patcog.2004.03.009]

[16] Zhang ML, Zhou ZH. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 2007, 40(7): 2038–2048. [doi: 10.1016/j.patcog.2006.12.019]

[17] Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: Proc. of the 15th Int'l Conf. on Neural Information Processing Systems: Natural and Synthetic. Vancouver: MIT Press, 2001. 681–687.

[18] Zhu Y, Kwok JT, Zhou ZH. Multi-label learning with global and local label correlation. IEEE Trans. on Knowledge and Data Engineering, 2018, 30(6): 1081–1094. [doi: 10.1109/TKDE.2017.2785795]

[19] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine Learning, 2011, 85(3): 333–359. [doi: 10.1007/s10994-011-5256-5]

[20] Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification. IEEE Trans. on Knowledge and Data Engineering, 2011, 23(7): 1079–1089. [doi: 10.1109/TKDE.2010.164]

[21] Lü SH, Chen YH, Jiang Y. Interaction-representation-based deep forest method in multi-label learning. Ruan Jian Xue Bao/Journal of Software, 2024, 35(4): 1934–1944 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6841.htm [doi: 10.13328/j.cnki.jos.006841]

[22] Liu JY, Jia XY. Multi-label classification algorithm based on association rule mining. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 2865–2878 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5341.htm [doi: 10.13328/j.cnki.jos.005341]

[23] Wang J, Yang Y, Mao JH, Huang ZH, Huang C, Xu W. CNN-RNN: A unified framework for multi-label image classification. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2285–2294. [doi: 10.1109/CVPR.2016.251]

[24] Yazici VO, Gonzalez-Garcia A, Ramisa A, Twardowski B, van de Weijer J. Orderless recurrent models for multi-label classification. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 13437–13446. [doi: 10.1109/CVPR42600.2020.01345]

[25] Chen ZM, Wei XS, Wang P, Guo YW. Multi-label image recognition with graph convolutional networks. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5172–5181. [doi: 10.1109/CVPR.2019.00532]

[26] Chen TS, Lin L, Chen RQ, Hui XL, Wu HF. Knowledge-guided multi-label few-shot learning for general image recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(3): 1371–1384. [doi: 10.1109/TPAMI.2020.3025814]

[27] Yeh CK, Wu WC, Ko WJ, Wang YCF. Learning deep latent space for multi-label classification. In: Proc. of the 31st AAAI Conf. on

Artificial Intelligence. San Francisco: AAAI, 2017. 2838–2844. [doi: 10.1609/aaai.v31i1.10769]

[28] Bai JW, Kong SF, Gomes C. Disentangled variational autoencoder based multi-label classification with covariance-aware multivariate probit model. In: Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence. Yokohama, 2021. 4313–4321.

[29] Zhan W, Zhang ML. Multi-label learning with label-specific features via clustering ensemble. In: Proc. of the 2017 IEEE Int'l Conf. on Data Science and Advanced Analytics. Tokyo: IEEE, 2017. 129–136. [doi: 10.1109/DSAA.2017.75]

[30] Zhang CY, Li ZS. Multi-label learning with label-specific features via weighting and label entropy guided clustering ensemble. Neurocomputing, 2021, 419: 59–69. [doi: 10.1016/j.neucom.2020.07.107]

[31] Zhang JJ, Fang M, Li X. Multi-label learning with discriminative features for each label. Neurocomputing, 2015, 154: 305–316. [doi: 10.1016/j.neucom.2014.11.062]

[32] Weng W, Lin YJ, Wu SX, Li YW, Kang Y. Multi-label learning based on label-specific features and local pairwise label correlation. Neurocomputing, 2018, 273: 385–394. [doi: 10.1016/j.neucom.2017.07.044]

[33] Zhang JD, Liu KY, Yang XB, Ju HR, Xu SP. Multi-label learning with relief-based label-specific feature selection. Applied Intelligence, 2023, 53(15): 18517–18530. [doi: 10.1007/s10489-022-04350-1]

[34] Mao JX, Hang JY, Zhang ML. Learning label-specific multiple local metrics for multi-label classification. In: Proc. of the 33rd Int'l Joint Conf. on Artificial Intelligence. Jeju, 2024. 4742–4750. [doi: 10.24963/ijcai.2024/524]

[35] Mao JX, Wang W, Zhang ML. Label specific multi-semantics metric learning for multi-label classification: Global consideration helps. In: Proc. of the 32nd Int'l Joint Conf. on Artificial Intelligence. Macao, 2023. 4055–4063. [doi: 10.24963/ijcai.2023/451]

[36] Mehravaran Z, Hamidzadeh J, Monsefi R. Feature selection based on correlation label and B-R belief function (FSCLBF) in multi-label data. Soft Computing, 2024, 28(2): 1445–1457. [doi: 10.1007/s00500-023-08341-3]

[37] Zhao TN, Zhang YJ, Miao DQ. Granular correlation-based label-specific feature augmentation for multi-label classification. Information Sciences, 2025, 689: 121473. [doi: 10.1016/j.ins.2024.121473]

[38] Hang JY, Zhang ML, Feng YH, Song XC. End-to-end probabilistic label-specific feature learning for multi-label classification. In: Proc. of the 36th AAAI Conf. on Artificial Intelligence. AAAI, 2022. 6847–6855. [doi: 10.1609/aaai.v36i6.20641]

[39] Weng W, Chen YN, Chen CL, Wu SX, Liu JH. Non-sparse label specific features selection for multi-label classification. Neurocomputing, 2020, 377: 85–94. [doi: 10.1016/j.neucom.2019.10.016]

[40] Zou YZ, Hu XG, Li PP, Ge YH. Learning shared and non-redundant label-specific features for partial multi-label classification. Information Sciences, 2024, 656: 119917. [doi: 10.1016/j.ins.2023.119917]

[41] Huang J, Li GR, Huang QM, Wu XD. Joint feature selection and classification for multilabel learning. IEEE Trans. on Cybernetics, 2018, 48(3): 876–889. [doi: 10.1109/TCYB.2017.2663838]

[42] Tan Y, Sun D, Shi Y, Gao LY, Gao QW, Lu YX. Bi-directional mapping for multi-label learning of label-specific features. Applied Intelligence, 2022, 52: 8147–8166. [doi: 10.1007/s10489-021-02868-4]

[43] Cheng ZW, Tan ZH. Multi-label learning for label-specific features using correlation information with missing label. Expert Systems with Applications, 2025, 269: 126491. [doi: 10.1016/j.eswa.2025.126491]

[44] Yu ZB, Zhang ML. Multi-label classification with label-specific feature generation: A wrapped approach. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(9): 5199–5210. [doi: 10.1109/TPAMI.2021.3070215]

[45] Hang JY, Zhang ML. Collaborative learning of label semantics and deep label-specific features for multi-label classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(12): 9860–9871. [doi: 10.1109/TPAMI.2021.3136592]

[46] Huang T, Jia B B, Zhang M L. Deep multi-dimensional classification with pairwise dimension-specific features. In: Proc. of the 33rd Int'l Joint Conf. on Artificial Intelligence. Jeju, 2024. 4183–4191. [doi: 10.24963/ijcai.2024/462]

[47] Zhao DW, Tan Y, Sun D, Gao QW, Lu YX, Zhu D. Multi-label learning of missing labels using label-specific features: An embedded packaging method. Applied Intelligence, 2024, 54(1): 791–814. [doi: 10.1007/s10489-023-05203-1]

[48] Lowe DG. Object recognition from local scale-invariant features. In: Proc. of the 7th IEEE Int'l Conf. on Computer Vision. Kerkyra: IEEE, 1999. 1150–1157. [doi: 10.1109/ICCV.1999.790410]

[49] Greenspan H, Belongie S, Goodman R, Perona P, Rakshit S, Anderson CH. Overcomplete steerable pyramid filters and rotation invariance. In: Proc. of the 1994 IEEE Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 1994. 222–228. [doi: 10.1109/CVPR.1994.323833]

[50] Cubuk ED, Zoph B, Mané D, Vasudevan V, Le QV. AutoAugment: Learning augmentation strategies from data. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 113–123. [doi: 10.1109/CVPR.2019.00020]

[51] Lee S, Cho S, Im S. DRANet: Disentangling representation and adaptation networks for unsupervised cross-domain adaptation. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 15247–15256. [doi: 10.1109/

CVPR46437.2021.01500]
[52] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929–1958.
[53] Huang G, Sun Y, Liu Z, Sedra D, Weinberger KQ. Deep networks with stochastic depth. In: Proc. of the 14th European Conf. on Computer Vision. Amsterdam: Springer, 2016. 646–661. [doi: 10.1007/978-3-319-46493-0_39]
[54] Achille A, Soatto S. Information dropout: Learning optimal representations through noisy computation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 40(12): 2897–2905. [doi: 10.1109/TPAMI.2017.2784440]
[55] Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: Proc. of the 3rd Int'l Conf. on Learning Representations. San Diego, 2015.
[56] Duan RJ, Chen YF, Niu DT, Yang Y, Qin AK, He Y. AdvDrop: Adversarial attack to DNNs by dropping information. In: Proc. of the 2021 IEEE/CVF Int'l Conf. on Computer Vision. Montreal: IEEE, 2021. 7486–7495. [doi: 10.1109/ICCV48922.2021.00741]
[57] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 1135–1144. [doi: 10.1145/2939672.2939778]
[58] Fong RC, Vedaldi A. Interpretable explanations of black boxes by meaningful perturbation. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision. Venice: IEEE, 2017. 3449–3457. [doi: 10.1109/ICCV.2017.371]
[59] Wilcoxon F. Individual comparisons by ranking methods. In: Kotz S, Johnson NL, eds. Breakthroughs in Statistics: Methodology and Distribution. New York: Springer, 1992. 196–202. [doi: 10.1007/978-1-4612-4380-9_16]

附中文参考文献

[21] 吕沈欢, 陈一赫, 姜远. 多标记学习中基于交互表示的深度森林方法. 软件学报, 2024, 35(4): 1934–1944. http://www.jos.org.cn/1000-9825/6841.htm [doi: 10.13328/j.cnki.jos.006841]
[22] 刘军煜, 贾修一. 一种利用关联规则挖掘的多标记分类算法. 软件学报, 2017, 28(11): 2865–2878. http://www.jos.org.cn/1000-9825/5341.htm [doi: 10.13328/j.cnki.jos.005341]

作者简介

杭均一, 博士生, 主要研究领域为机器学习, 多标记分类.
张敏灵, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为机器学习, 数据挖掘.