

集成特征注意力和残差连接的偏标签回归算法

吴海峰^{1,2}, 陶丽青¹, 程玉胜^{1,2*}

(1. 安庆师范大学 计算机与信息学院, 安徽 安庆 246133; 2. 智能感知与计算安徽省高校重点实验室(安庆师范大学), 安徽 安庆 246133)
(* 通信作者电子邮箱 chengyusheng@163.com)

摘要: 偏标签回归(PLR)弥补了偏标签学习(PLL)仅聚焦于分类任务的局限。针对现有的PLR算法忽略实例特征的特性差异的问题, 提出一种集成特征注意力和残差连接的偏标签回归算法(PLR-FARC)。首先, 通过标签增强技术将真实数据集的标签扩充为一组实值候选标签; 其次, 借助注意力机制自动生成每个特征对标签的贡献度; 再次, 引入残差连接以减少特征在传递过程中的信息丢失, 从而维持特征的完整性; 最后, 分别基于IDent (IDentification method)和PIDent (Progressive IDentification method)计算预测损失。在Abalone、Airfoil、Concrete、Cpu-act、Housing和Power-plant数据集上的实验结果表明, 相较于IDent和PIDent, PLR-FARC的平均绝对误差(MAE)分别平均降低了2.15%、38.38%、8.86%、4.19%、15.71%和15.55%, 均方误差(MSE)分别平均降低了9.35%、71.32%、23.10%、20.17%、27.22%和9.46%。可见, 所提算法是可行且有效的。

关键词: 偏标签学习; 偏标签回归; 候选标签; 注意力机制; 残差连接

中图分类号: TP311.5 **文献标志码:** A

Partial label regression algorithm integrating feature attention and residual connection

WU Haifeng^{1,2}, TAO Liqing¹, CHENG Yusheng^{1,2*}

(1. School of Computer and Information, Anqing Normal University, Anqing Anhui 246133, China;
2. University Key Laboratory of Intelligent Perception and Computing of Anhui Province (Anqing Normal University), Anqing Anhui 246133, China)

Abstract: Partial Label Regression (PLR) complements the current situation that Partial Label Learning (PLL) only focuses on classification tasks. To address the problem that the existing PLR algorithms ignore characteristic differences between instance features, a Partial Label Regression algorithm integrating Feature Attention and Residual Connection (PLR-FARC) was proposed. Firstly, labels of real datasets were expanded into a set of real-value candidate labels by label enhancement technique. Secondly, the attention mechanism was employed to generate contribution of individual features to labels automatically. Thirdly, the residual connection was introduced to reduce information loss and maintain feature integrity during feature transmission. Finally, prediction loss was calculated based on IDent (IDentification method) and PIDent (Progressive IDentification method), respectively. Experimental results on Abalone, Airfoil, Concrete, Cpu-act, Housing and Power-plant datasets show that compared to IDent and PIDent, PLR-FARC has the Mean Absolute Error (MAE) reduced by 2.15%, 38.38%, 8.86%, 4.19%, 15.71% and 15.55%, averagely and respectively, and the Mean Squared Error (MSE) reduced by 9.35%, 71.32%, 23.10%, 20.17%, 27.22% and 9.46%, averagely and respectively. It can be seen that the proposed algorithm is feasible and effective.

Key words: Partial Label Learning (PLL); Partial Label Regression (PLR); candidate label; attention mechanism; residual connection

0 引言

偏标签学习(Partial Label Learning, PLL)是一种新兴的弱监督学习框架,旨在解决标签噪声问题^[1]。在PLL中,实例与候选标签集密切相关。候选标签集中有且只有一个真实标签,且该真实标签无法被算法直接获取^[2]。此外,每个候选标签被认为是真实标签的概率相等,换言之,对于研究者,样本的真实标签是潜在未知的。PLL擅于处理具有标签歧义的实例样本,在多个领域发挥重要作用,例如多媒体内容

分析^[3]、生态学信息^[4]、图像自动标注^[5]和网络挖掘^[6]等。

在PLL中,候选标签集中噪声标签会对模型训练产生负面影响,因此,如何有效消除歧义的噪声标签是PLL的关键^[7]。为解决噪声问题,研究人员从技术方法策略研究和理论分析研究2个方面开展工作。技术方法策略的研究包括基于消歧的偏标签学习方法和基于无消歧的偏标签学习方法。基于平均的消歧方法,在训练中一视同仁地对待每一个候选标签,将候选标签的平均建模输出值作为预测值。例如Gong等^[8]将LMNN(Large Margin Nearest Neighbor)框架应用

收稿日期:2024-07-18;修回日期:2024-11-12;录用日期:2024-11-13。

基金项目:安徽省自然科学基金资助项目(2108085MF216);安徽省高校自然科学基金项目(2024AH040175, 2024AH051099)。

作者简介:吴海峰(1982—),男,安徽安庆人,教授,硕士,主要研究方向:机器学习、数据挖掘; 陶丽青(1999—),女,安徽铜陵人,硕士研究生,主要研究方向:偏标签学习、数据挖掘; 程玉胜(1969—),男,安徽安庆人,教授,博士,主要研究方向:粗糙集、半监督学习、机器学习、数据挖掘。

于偏标签学习,通过重新组织数据的底层结构学习新的度量标准。然而,在基于平均的消歧方法中,由于所有标签被平等对待,噪声标签极易对模型产生误导。基于识别的消歧方法,将候选标签集中的真实标签作为潜在变量,通过迭代优化候选标签的置信度识别真实标签。例如,Gu等^[9]提出了DEER(Distribution Divergence-based Graph Contrast),通过图增强技术推断隐藏空间中的潜在语义分布,并基于分布差异识别正样本对。基于识别的消歧方法考虑了候选标签间的差异,但在识别的过程中依赖于置信度的估计精度,并引入了额外的复杂性。基于无消歧的偏标签学习方法直接对整个候选标签集的分布进行建模,学习得到最佳预测,减少了消歧过程中潜在的不确定性。例如,Lin等^[10]结合特征空间和标签空间,利用纠错输出码(Error-Correcting Output Codes, ECOC)框架进行偏标签研究。理论分析层面的研究聚焦于偏标签学习方法的性能保证和有效性验证。例如,Liu等^[11]提出小模糊度条件,为噪声标签提供了一种理论上的控制机制,确保模型在候选标签集中存在噪声的情况下仍能准确预测;文献[12]中的统计一致性,用于验证在偏标签上计算的经验风险与在普通标签上计算的经验风险是否一致。传统的偏标签学习基于同一个前提假设^[13],即实例的候选标签是离散的,致力于离散标签的分类处理,未涉及回归任务。

偏标签回归(Partial Label Regression, PLR)^[14]是PLL的变体和延伸,在理论上填补了PLL在回归任务中的空缺,丰富了偏标签学习的研究。在PLR中,实例被标注为一组连续实值标签,且只有一个真实标签。PLL致力于离散的分类任务,PLR关注连续实值的回归预测。在图1中,PLL将图片标注为(猴子,松鼠,狐狸),猴子、松鼠和狐狸是离散的分类标签,“猴子”为图片的真实标签;PLR将图片标注为(3,5,7),3、5、7是连续的实值标签,“3”为猴子的真实年龄。在现实场景中,具有一组连续实值候选标签的偏标签回归任务与离散标签的偏标签分类任务一样,都是普遍存在的问题。因此,PLR具有良好的研究前景,学习有效的PLR模型无论是在理论上还是在现实场景中,都具有重要的研究意义。

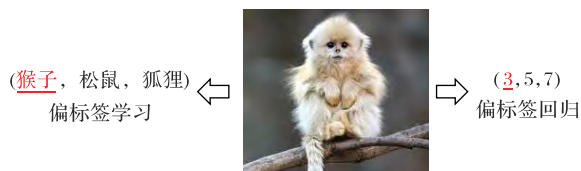


图1 PLL与PLR示例

Fig. 1 Instance of PLL and PLR

文献[14]率先开展了偏标签回归任务的尝试,关注候选标签的特性差异,着重于回归模型预测损失设计,提出Baseline、IDent(IDentification method)和PIDent(Progressive IDentification method)。然而,Baseline、IDent和PIDent均只关注候选标签的特性差异,忽略了特征的特性差异问题。因此,本文提出集成特征注意力和残差连接的偏标签回归算法(Partial Label Regression algorithm integrating Feature Attention and Residual Connection, PLR-FARC)。

本文的主要工作包括3个方面:

1)关注实例特征的特性差异,引入特征注意力机制评估特征的贡献度。借助注意力机制动态分配特征权重,精准捕

捉重要特征,进而提高偏标签回归模型的鲁棒性和泛化能力。

2)引入残差连接,确保特征有效传递,解决特征传递过程中的信息退化问题,从而增强偏标签回归模型的稳定性。

3)从特征的角度研究偏标签回归,提出PLR-FARC。将特征注意力与残差连接相结合,有效解决了现有算法忽略特征特性差异的问题,提升了偏标签回归模型的预测性能。

1 相关工作

Cheng等^[14]首次提出偏标签回归概念,并基于候选标签的特性差异,提出Baseline、IDent和PIDent这3种用于评估偏标签回归模型预测性能的算法。Baseline把所有候选标签产生损失的平均值作为预测损失,由于它未考虑候选标签的特性差异,真实标签产生的损失易被噪声标签产生的损失掩盖。为了能关注候选标签的特性差异,减小噪声标签对真实标签的影响,IDent被提出。IDent把最小损失的候选标签当作真实标签,并将已识别噪声标签中的最小损失作为偏标签回归模型的预测损失。Baseline平等对待所有候选标签,而IDent过度关注单个标签。为弥补二者的不足,Cheng等^[14]最终提出了PIDent。PIDent在IDent的基础上引入权重设置,将逐步更新的损失视为偏标签回归模型的预测损失,既考虑了候选标签的特性差异,又避免了过分关注单一标签。

注意力机制在机器学习和深度学习中发挥重要作用。Yuan等^[15]将注意力机制引入复杂工业过程中的质量预测,在不同尺度的特征图中并行设计通道注意力机制,获取注意力权重以提高评估预测性能;Huang等^[16]将注意力机制应用于图像修复领域,提出一种稀疏自注意力转换器。

残差连接将输出视为输入与一个非线性变换输入的线性叠加^[17]。Russakovsky等^[18]首次将残差连接用于构建深层次残差网络(Residual Network, ResNet),极大地改善了网络的训练效果。随后,残差连接在大模型中被广泛应用,如Transformer^[19]和BERT(Bidirectional Encoder Representations from Transformers)^[20]等。

本文算法PLR-FARC与现有偏标签回归算法IDent和PIDent密切相关。PLR-FARC不仅关注候选标签的特性差异,也考虑了偏标签回归实例特征的特性差异,既引入特征注意力机制评估特征的贡献度,同时引入残差连接确保特征信息的有效传递。

2 PLR-FARC

本文从特征的角度研究偏标签回归,提出PLR-FARC,能有效地从具有一组实值候选标签的数据中训练偏标签回归模型。如图2,PLR-FARC包括标签增强模块、多层感知机(Multi-Layer Perception, MLP)模块、特征贡献度模块以及预测模块。

2.1 标签增强以生成候选标签

为得到偏标签回归数据集,借助均匀分布策略^[21]构造实例的弱监督信息,将真实数据集的实值标签转换为一组连续的实值候选标签。假设所有样本实例的实值标签服从均匀分布,它的概率密度:

$$p(y_{\min}, y_{\max}) = 1/(y_{\max} - y_{\min}) \quad (1)$$

其中: $y_{\max}$ 表示实值标签空间中的最大值; $y_{\min}$ 表示实值标签

空间中的最小值。当样本实例的候选标签数为 $\max_set$ 时, 从 $(y_{\min}, y_{\max})$ 区间范围内随机生成 $\max_set - 1$ 个噪声标签。噪声标签与实例的真实标签组合, 共同构成实例的候选标签集 G 。

基于式 (1), 得到适用于偏标签回归的数据集 $\hat{D} = \{(\mathbf{x}_i, G_i) | 1 \leq i \leq n, G_i \subseteq Y\}$ 。 $\mathbf{x}_i = \{x_i^1, x_i^2, \dots, x_i^l\}$ 表示实例

$\mathbf{x}_i \in X$ 的特征空间; $Y = \{y_1, y_2, \dots, y_n\}$ 表示实例 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 的真实标签。此外, 假设 PLR-FARC 满足小模糊度:

$$\gamma = \sup_{(\mathbf{x}, y) \sim p(\mathbf{x}, y), (\mathbf{x}, G) \sim p(\mathbf{x}, G), y' \neq y, y' \in Y} p(y' \in G) \quad (2)$$

其中: γ 表示候选标签集中噪声标签的最大概率; y' 为噪声标签。若 $\gamma = 1$, 则无法区分真实标签和噪声标签。

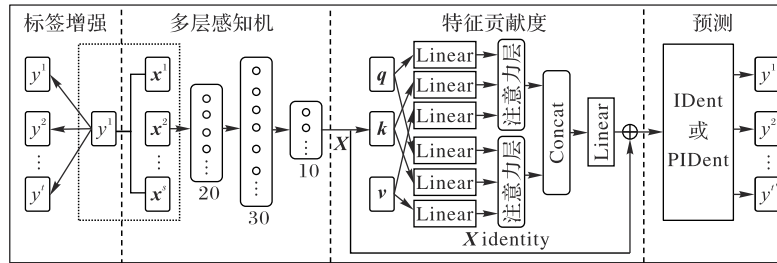


图2 PLR-FARC的框架

Fig. 2 Framework of PLR-FARC

2.2 基于注意力机制计算特征贡献度

一方面, 由于实例特征的特性差异, 不同维度的特征对模型训练的贡献度不同; 另一方面, 注意力机制能计算特征间的相关性从而判断特征的重要性^[22]。因此, PLR-FARC 将注意力机制引入偏标签回归, 构造基于多头注意力机制的特征贡献度模块。该模块旨在通过计算学习得到一个非线性函数以自动计算特征的贡献度, 并构造新的带有特征贡献度信息的特征向量。

自注意力机制能在序列数据中动态地学习和分配注意力权重^[23]。自注意力机制将自身特征与其他特征关联, 完成特征间的相关性计算, 得到特征注意力分数, 最终实现特征的注意力权重分配。自注意力机制的特征注意力权重计算公式如下:

$$\text{Atten}(\mathbf{q}, \mathbf{k}, \mathbf{v}) = \text{Softmax}\left(\frac{\mathbf{qk}^T}{\sqrt{d}}\right) \cdot \mathbf{v} \quad (3)$$

其中: $\mathbf{q}$ 、 $\mathbf{k}$ 和 $\mathbf{v}$ 分别表示查询向量、键向量和值向量; d 表示特征的维度。首先, 自注意力机制将输入的序列数据线性转换为 $\mathbf{q}$ 、 $\mathbf{k}$ 和 $\mathbf{v}$; 其次, 借助 $\mathbf{qk}^T$ 计算特征的注意力分数, $\sqrt{d}$ 防止出现特征维度过大而导致梯度消失的问题; 最后, 将归一化后的注意力分数与 $\mathbf{v}$ 点乘, 得到最终的特征注意力权重。

PLR-FARC 采用多头自注意力机制计算实例特征的特征注意力权重, 最终实现特征的贡献度分配。多头自注意力机制将自注意力机制扩展至多维空间, 引入多组注意力权重, 从而多角度地捕捉序列数据中的信息^[24]。多头注意力机制将输入序列进行多个不同的线性转换, 针对每个不同的线性转换分别计算注意力权重, 再将多个注意力权重拼接, 得到最终的输出序列。多头注意力机制的计算如下:

$$\text{MultiHead}(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^h) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) \mathbf{W}^o \quad (4)$$

$$\text{head}_t = \text{Atten}(\mathbf{W}_q^t \mathbf{x}^t, \mathbf{W}_k^t \mathbf{x}^t, \mathbf{W}_v^t \mathbf{x}^t); 1 \leq t \leq h \quad (5)$$

其中: $\mathbf{W}^o$ 、 $\mathbf{W}_q^t$ 、 $\mathbf{W}_k^t$ 、 $\mathbf{W}_v^t$ 是参数向量; t 是头数。

2.3 残差连接

随着网络层数增加, 在传递的过程中极易丢失特征的细粒度信息。为了维持特征的完整性, PLR-FARC 引入残差连接。残差连接的核心思想是引入一种特殊的跳跃连接, 在特

定条件下, 允许输入直接跳过某些网络层, 并与后续网络层直接相连, 有助于在网络层中保留输入的原始信息^[25]。

在 PLR-FARC 中, 输入特征 $\mathbf{x}$ 经过 MLP 处理后, 进入特征贡献模块计算特征的贡献, 由于残差连接, 最终输出特征:

$$\mathbf{x}' = \text{MultiHead}(\text{MLP}(\mathbf{x})) + \text{MLP}(\mathbf{x}) \quad (6)$$

其中 $\text{MLP}(\mathbf{x})$ 表示经过 MLP 处理的输入特征。

2.4 基于 IDent 和 PIDent 的回归损失计算

回归任务的目标是获得回归模型, 使期望风险^[26]最低:

$$L(f) = E_{p(\mathbf{x}, y)}[\text{loss}(f(\mathbf{x}), y)] \quad (7)$$

其中: $E_{p(\mathbf{x}, y)}$ 为概率分布密度 $p(\mathbf{x}, y)$ 数据的期望; $\text{loss}(\cdot)$ 是传统的损失函数, 如平均绝对误差 (Mean Absolute Error, MAE)^[27] 和均方误差 (Mean Squared Error, MSE)^[28]。由于回归模型是通过训练数据集中部分实例得到的, 无法获得可靠 $p(\mathbf{x}, y)$ 。此时, 常见策略是采用经验风险:

$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^n \text{loss}(f(\mathbf{x}_i), y_i) \quad (8)$$

PLR-FARC 基于 IDent 和 PIDent 计算模型的预测损失。

IDent 将候选标签集中的最小损失视为真实标签, 同时, 将已识别的噪声标签中最小损失作为模型的预测损失。因此, 基于 IDent 的偏标签回归模型预测损失表示为:

$$\text{Loss}_{\text{ID}}(f(\mathbf{x}), G) = \min_{y \in G} \text{loss}(f(\mathbf{x}), y) \quad (9)$$

PIDent 在 IDent 的基础上加入权重设置, 将候选标签的加权损失作为模型的预测损失。因此, 基于 PIDent 的偏标签回归模型预测损失表示为:

$$\text{Loss}_{\text{PID}}(f(\mathbf{x}), y) = \sum_{y \in G} w(\mathbf{x}, y) \text{loss}(f(\mathbf{x}), y) \quad (10)$$

其中: $w(\mathbf{x}, y)$ 是权重参数, 反映候选标签的重要性。候选标签的重要性除了借助权重表现之外, 候选标签产生的损失也能反映候选标签的重要性程度, 即候选标签产生的损失越小, 权重越大; 反之, 权重越小。因此, 权重函数为:

$$w(\mathbf{x}, y) = \begin{cases} \frac{\exp(\alpha * \text{Loss}_{\text{PID}}(f(\mathbf{x}), y)^{-\beta})}{\sum_{y \in G} \exp(\alpha * \text{Loss}_{\text{PID}}(f(\mathbf{x}), y)^{-\beta})}, & y \in G \\ 0, & \text{其他} \end{cases} \quad (11)$$

其中: α 和 β 是超参数。 $\alpha * \text{Loss}_{\text{PID}}(f(\mathbf{x}), y)^{-\beta}$ 意味着较小损失

的候选标签具有较大的权重,再依靠 Softmax 将权重归一化;同时,将非实例候选标签的权重置为 0。由于候选标签具有潜在未知性,每个候选标签都可能是实例的真实标签,满足:

$$w(\mathbf{x}, y) \geq 0 \text{ 且 } \sum_{y \in G} w(\mathbf{x}, y) = 1 \quad (12)$$

基于式(8),IDent 和 PIDent 经验风险可分别表示为:

$$\hat{L}_{\text{ID}}(f) = \frac{1}{n} \sum_{i=1}^n \text{Loss}_{\text{ID}}(f(\mathbf{x}_i), \mathbf{G}) \quad (13)$$

$$\hat{L}_{\text{PID}}(f) = \frac{1}{n} \sum_{i=1}^n \text{Loss}_{\text{PID}}(f(\mathbf{x}_i), \mathbf{G}) \quad (14)$$

最后,通过最小化经验风险,迭代训练,得到有效的偏标签回归模型。

集成特征注意力和残差连接的偏标签回归算法 PLR-FARC,算法流程如下。

输入 特征空间 $\mathbf{X}$, 标签空间 $\mathbf{Y}$, 模型 f , 权重参数 $w(\mathbf{x}, y)$, 偏标签回归损失函数 L , 遍历次数 $E_{\max}$, 迭代次数 $I_{\max}$ 。

输出 模型 f 。

- 1) 根据式(1)生成候选标签集 $\mathbf{G}$, 得到偏标签回归数据集 $\hat{\mathbf{D}} = \{(\mathbf{x}_i, \mathbf{G}_i) | 1 \leq i \leq n, \mathbf{G}_i \subseteq \mathbf{Y}\}$
- 2) 根据式(5)分配特征权重, 得到新的特征向量 $\mathbf{X}'$
- 3) 初始化模型 f , 权重参数 $w(\mathbf{x}, y) = 1/|\mathbf{G}|$
- 4) for $e = 1, 2, \dots, E_{\max}$
- 5) 随机打乱 $\hat{\mathbf{D}}$
- 6) for $i = 1, 2, \dots, I_{\max}$
- 7) 从 $\hat{\mathbf{D}}$ 中提取最小批次 $\hat{\mathbf{D}}_{\min}$
- 8) if L 是 IDent, 根据式(13)更新模型 f
- 9) if L 是 PIDent, 根据式(14)更新模型 f
- 10) end for
- 11) 根据式(11)更新权重参数 $w(\mathbf{x}, y)$
- 12) end for

3 实验与结果分析

3.1 数据集和评价指标

为验证 PLR-FARC 的有效性,本文选择 UCI (University of California Irvine) 机器学习库中 6 个广泛使用的基准回归数据集,包括 Abalone、Airfoil、Concrete、Cpu-act、Housing 和 Power-plant,数据集的详细信息如表 1 所示。首先,对实例特征向量预处理,包括对具有固定范围的维度进行最大最小归一化处理、对离散变量进行 one-hot 编码,以及对连续实值维度进行均值为 0、标准差为 1 的标准化处理;其次,为构建适用于偏标签回归的数据集,本文使用标签增强技术扩充标签空间,生成一组连续实值的候选标签;最后,将预处理后的数据集随机划分为训练集 (60%)、验证集 (20%) 和测试集 (20%)。

表 1 实验中使用的基准数据集

Tab. 1 Benchmark datasets used in experiments

数据集	特征数	样本数		
		训练集	验证集	测试集
Abalone	8	2 507	835	835
Airfoil	5	903	300	300
Concrete	8	618	206	206
Cpu-act	21	4 916	1 638	1 638
Housing	13	304	101	101
Power-plant	4	5 742	1 913	1 913

本文采用回归学习中常见的评价指标 MAE^[27] 和 MSE^[28], 评估 PLR-FARC 与 8 个对比算法的性能。MAE 和 MSE 的计算方法分别为:

$$V_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i' - y_i| \quad (15)$$

$$V_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i' - y_i)^2 \quad (16)$$

其中: y_i' 为预测标签; y_i 为真实标签; n 表示样本数。MAE 和 MSE 值越小, 算法性能越好。

3.2 对比算法和参数设置

在 PLR-FARC 中, MLP 是一个 5 层 (s -20-30-10-1) 神经网络, 使用 GELU (Gaussian Error Linear Unit) 激活函数和 Adam (Adaptive moment estimation) 优化方法, 学习率 $r \in \{0.001, 0.010\}$, 批量大小和训练轮数分别设置为 256 和 1 000。

实验设置 4 种候选标签集规模, $|\mathbf{G}| = 2, |\mathbf{G}| = 4, |\mathbf{G}| = 8$ 和 $|\mathbf{G}| = 16$ 分别表示候选标签集中的标签数为 2、4、8 和 16。在基准数据集上, 将 PLR-FARC 与文献 [14] 中的 6 种 Baseline、IDent 以及 PIDent 对比, 参数设置如下:

1) Baseline。包括 AVG 和 AVGL。

AVGV: 对于每个样本实例, 将候选标签集中所有候选标签的平均值作为真实标签, 并采用最小化传统损失函数的经验风险以训练回归模型。

AVGL: 对于每个样本实例, 平等地看待候选标签集中的每一个候选标签, 并将所有候选标签的平均损失作为回归模型的预测损失。

AVGV 和 AVGL 可以配置任意的回归损失函数, 本文选取 MSE、MAE 和 Huber 损失。因此, 基于 AVGV 和 AVGL 的算法包括 AVGL-MSE、AVGL-MAE、AVGL-Huber、AVGV-MSE、AVGV-MAE 和 AVGV-Huber。

2) IDent。对于每个实例样本, 将候选标签集中产生最小损失的候选标签视为真实标签; 同时, 将已识别的假阳性标签中的最小损失作为回归模型的预测损失。IDent 将 MSE 作为自身的回归损失函数。

3) PIDent。对于每个实例样本, 将候选标签的加权损失作为回归模型的预测损失。PIDent 将 MSE 作为自身的回归损失函数, $\alpha = \{10, 100, 500, 1\,000, 10\,000\}$, $\beta = 0.5$ 。

4) PLR-FARC。关注实例特征的特性差异。同时, 在传递过程中维持特征的完整性, 减少细粒度信息的丢失。由于 PLR-FARC 不依赖于特定的偏标签回归损失计算方法, 本文分别选取 IDent 和 PIDent。

为便于区分, 本文将基于 IDent 的 PLR-FARC 简称为 FARCID (PLR-FARC based on IDent), 将基于 PIDent 的 PLR-FARC 简称为 FARCPID (PLR-FARC based on PIDent)。

3.3 实验结果和分析

表 2、3 分别展示了 PLR-FARC 与对比算法在 MAE 和 MSE 评价指标上的性能, 最优结果加粗表示, 下划线表示次优结果。实验结果表明, PLR-FARC 包含了多数的最优结果和次优结果。相较于 8 种对比算法, PLR-FARC 具有显著的性能优势, 充分证明了该算法的有效性。

表2 各算法的MAE性能对比

Tab. 2 Performance comparison of various algorithms in MAE

数据集	G	AVGL-MSE	AVGL-MAE	AVGL-Huber	AVGV-MSE	AVGV-MAE	AVGV-Hube	IDent	PIDent	PLR-FARC	
										FARCID	FARCPID
Abalone	2	2.15	1.48	1.48	2.15	2.15	1.69	<u>1.45</u>	1.49	<u>1.45</u>	1.42
	4	2.75	1.56	1.64	2.75	2.75	2.77	1.48	1.52	<u>1.47</u>	1.45
	8	3.10	1.95	2.07	3.10	3.10	3.46	<u>1.51</u>	<u>1.51</u>	1.52	1.50
	16	3.16	2.88	2.87	3.16	3.16	3.62	1.55	<u>1.54</u>	<u>1.54</u>	1.48
Airfoil	2	3.8	3.19	3.21	3.80	3.41	3.66	2.21	2.15	1.44	<u>1.50</u>
	4	4.16	3.26	3.30	4.16	3.97	3.99	2.26	2.51	<u>1.57</u>	1.53
	8	4.80	3.65	3.74	4.80	4.66	5.02	2.99	3.29	1.52	<u>1.84</u>
	16	5.40	4.33	4.44	5.40	5.18	5.40	4.07	4.33	2.48	<u>2.87</u>
Concrete	2	8.17	5.14	5.10	8.17	7.79	7.79	4.78	4.77	4.47	<u>4.59</u>
	4	8.70	5.36	5.57	8.69	8.77	8.77	5.13	5.42	5.30	<u>5.28</u>
	8	10.87	7.71	7.82	10.83	10.53	10.53	5.96	6.51	5.53	<u>5.57</u>
	16	11.97	9.26	9.26	11.97	12.45	12.45	7.79	7.93	5.85	<u>7.51</u>
Cpu-act	2	11.92	2.06	2.15	11.70	8.72	8.58	1.89	<u>1.81</u>	1.78	<u>1.81</u>
	4	19.27	3.65	3.92	18.65	19.34	18.28	1.91	1.87	<u>1.84</u>	1.83
	8	23.81	15.29	14.98	23.81	27.95	28.20	1.93	<u>1.92</u>	1.82	1.95
	16	29.08	23.73	23.17	29.07	30.57	30.65	2.07	2.06	1.91	<u>1.97</u>
Housing	2	4.15	2.99	3.33	4.15	4.97	5.06	<u>2.62</u>	3.22	2.16	2.64
	4	4.81	3.15	3.42	4.81	5.19	5.26	3.05	3.07	2.78	<u>2.81</u>
	8	5.84	4.29	4.56	5.84	6.52	6.47	3.32	4.93	2.80	<u>3.20</u>
	16	6.61	5.06	5.02	6.61	7.06	7.48	5.06	<u>4.52</u>	4.88	3.58
Power-plant	2	6.45	3.88	3.86	6.45	5.37	5.71	3.70	3.70	<u>3.10</u>	3.04
	4	8.57	4.28	4.48	8.57	8.50	8.35	3.68	3.68	<u>3.18</u>	3.08
	8	10.35	5.41	5.89	10.35	10.92	11.01	3.69	3.69	<u>3.39</u>	3.13
	16	11.96	8.77	8.44	11.96	12.64	13.10	<u>3.72</u>	3.73	5.70	3.24

表3 各算法的MSE性能对比

Tab. 3 Performance comparison of various algorithms in MSE

数据集	G	AVGL-MSE	AVGL-MAE	AVGL-Huber	AVGV-MSE	AVGV-MAE	AVGV-Huber	IDent	PIDent	PLR-FARC	
										FARCID	FARCPID
Abalone	2	9.70	4.66	4.68	9.72	7.36	7.52	4.62	4.55	<u>4.06</u>	4.04
	4	14.42	5.04	5.22	14.08	14.08	14.04	4.66	4.58	<u>4.24</u>	4.08
	8	20.63	7.76	7.94	22.68	22.68	21.28	4.70	4.71	<u>4.47</u>	4.22
	16	25.11	13.77	14.03	26.71	26.71	26.61	4.90	4.90	4.65	<u>4.73</u>
Airfoil	2	23.72	16.66	16.17	23.73	23.27	22.99	15.58	14.99	3.63	<u>4.51</u>
	4	30.79	18.47	18.56	30.83	30.67	30.22	16.23	16.10	4.05	<u>6.48</u>
	8	37.19	24.67	24.40	37.16	39.24	38.19	17.81	17.86	5.96	<u>10.27</u>
	16	42.90	31.43	30.81	43.00	45.38	44.19	23.41	24.11	7.46	<u>19.94</u>
Concrete	2	108.08	46.66	48.62	108.23	107.46	106.00	42.14	40.48	<u>38.91</u>	32.73
	4	151.05	75.49	80.38	151.05	172.01	167.53	45.61	<u>44.87</u>	45.94	34.09
	8	195.40	116.18	118.65	195.32	216.01	213.59	72.09	<u>63.79</u>	57.79	75.78
	16	239.26	186.94	183.64	239.64	268.56	259.05	114.35	112.02	67.38	<u>96.18</u>
Cpu-act	2	246.38	9.84	9.62	245.68	172.38	169.26	6.64	6.91	5.66	<u>6.53</u>
	4	461.69	33.15	34.58	461.05	525.78	520.64	7.00	7.38	5.74	<u>6.69</u>
	8	730.38	335.04	333.92	730.93	858.55	852.76	7.48	7.90	5.97	<u>6.95</u>
	16	972.73	738.42	735.89	996.21	1107.80	1106.00	9.48	9.11	6.58	<u>7.07</u>
Housing	2	33.82	18.69	17.37	33.82	32.79	34.03	16.18	15.55	<u>14.79</u>	13.32
	4	48.82	26.78	26.82	48.80	51.87	53.67	21.59	<u>20.53</u>	22.49	13.74
	8	60.87	38.18	39.19	60.90	64.02	62.23	29.82	26.87	<u>22.54</u>	21.42
	16	75.30	52.24	53.06	75.29	84.18	77.00	39.88	41.09	23.44	<u>37.42</u>
Power-plant	2	64.99	23.67	23.43	64.99	42.56	42.68	21.09	21.07	<u>19.81</u>	18.47
	4	104.61	29.24	29.47	104.61	98.36	98.41	21.29	21.28	<u>19.50</u>	19.15
	8	153.07	153.07	48.99	153.07	167.96	166.21	21.21	21.36	<u>20.16</u>	19.24
	16	204.32	204.32	105.45	204.31	225.97	221.17	21.31	21.34	<u>20.29</u>	19.98

具体地,当评价指标为MAE时,在 Abalone 和 Power-plant 上, FARCPID 性能更优;在 Airfoil、Concrete、Cpu-act 和 Housing 上, FARCID 性能更佳;在数据集 Abalone、Airfoil、Concrete、Cpu-act、Housing 和 Power-plant 上, PLR-FARC 性能

分别平均提升了 2.15%、38.38%、8.86%、4.19%、15.71% 以及 15.55%。当评价指标为 MSE 时,在 Abalone、Housing 和 Power-plant 上, FARCPID 性能更优;在 Airfoil 和 Cpu-act 上, FARCID 性能更佳;在 Concrete 上, FARCID 与 FARCPID 性能相当;在数据集 Abalone、Airfoil、Concrete、Cpu-act、Housing 和 Power-plant 上, PLR-FARC 性能分别平均提升了 9.35%、71.32%、23.10%、20.17%、27.22% 以及 9.46%。

PLR-FARC 的优异表现源于该算法充分利用了实例特征的差异性,同时,通过残差连接有效减少了信息传递过程中的损失。随着候选标签集规模的扩大, PLR-FARC 性能下降,这是因为在候选标签集中噪声标签增多,偏标签回归任务的难度增大。

3.4 消融实验

为进一步验证关注实例特征的特性差异和维持特征传递过程中的完整性是否有助于提升偏标签回归算法的性能,本文基于 MSE 评价指标和 Abalone 数据集进行消融实验^[29]。表 4 和图 3 展示消融实验结果。其中, PLR-FA (Partial Label Regression algorithm based on Feature Attention) 仅包含注意力模块, PLR-RC (Partial Label Regression Algorithm based on Residual Connection) 仅包含残差模块, PLR-FARC 则同时包含注意力模块和残差模块。由于 PLR-FA、PLR-RC 和 PLR-FARC 不依赖于特定的偏标签回归损失方法,本文选择 2 种不同的损失计算方法,即 IDent 和 PIDent。

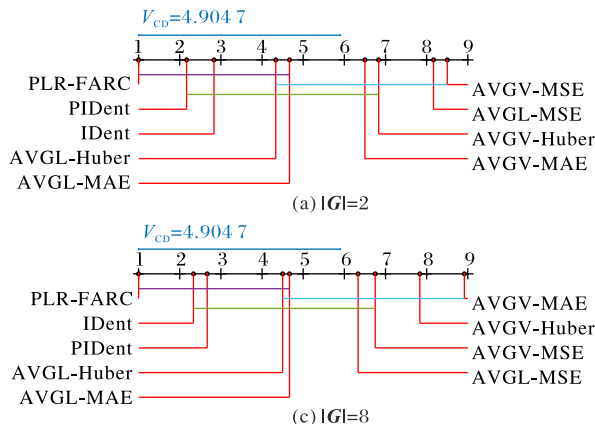
表 4 Abalone 数据集上消融实验的 MSE 性能对比

Tab. 4 Performance comparison of ablation experiments on

Abalone dataset in MSE

注意力 模块	残差 模块	算法	不同 G 时的MSE				
			2	4	8	16	
√	×	PLR-FA	+IDent	4.35	4.48	4.49	4.86
			+PIDent	4.27	4.41	4.68	4.77
×	√	PLR-RC	+IDent	4.28	4.34	4.63	4.82
			+PIDent	4.37	4.38	5.06	5.68
√	√	PLR-FARC	+IDent	4.06	4.24	4.47	4.65
			+PIDent	4.04	4.08	4.22	4.73

如表 4 所示, PLR-FARC 占据最优结果(加粗)和次优结



果(下划线),证明了集成注意力和残差连接能有效提升偏标签回归算法性能。

图 3 直观地展示了 PLR-FA、PLR-RC 和 PLR-FARC 性能的变化趋势。随着候选标签集规模增大, 3 种算法的性能均降低, 进一步说明了噪声标签的增加会降低算法的性能。

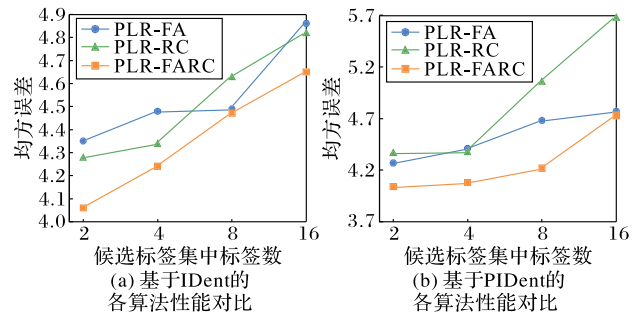


图 3 基于 IDent 和 PIDent 的消融实验结果

Fig. 3 Ablation experimental results based on IDent and PIDent

3.5 统计假设检验

为更清晰地展示 PLR-FARC 在 6 个数据集上的综合性能, 本文基于 MSE 评价指标, 采用显著性水平为 5% 的 Nemenyi 检验^[30]进行统计假设检验。若 2 个算法的平均排序差大于临界差 (Critical Difference, CD), 则说明两算法间具有显著差异, 否则无显著性差异。CD 值的计算方法为:

$$V_{CD} = q_{\alpha} \sqrt{\frac{K(K+1)}{6N}} \quad (17)$$

其中: 显著性水平 $\alpha = 0.05$; $K = 9$; $N = 6$; $q_{\alpha} = 3.102$; $V_{CD} = 4.9047$ 。

图 4 分别展示了 PLR-FARC 在 $|G| = 2$ 、 $|G| = 4$ 、 $|G| = 8$ 和 $|G| = 16$ 时与其他 8 种算法的性能比较, 各算法性能从左向右依次降低。PLR-FARC 性能最优, 与 AVGL-MSE、AVGV-MSE、AVGV-MAE 和 AVGV-Huber 相比有显著性差异, 与 IDent、PIDent、AVGL-MAE 和 AVGL-Huber 相比无显著性差异。

综上, PLR-FARC 具有更优性能。

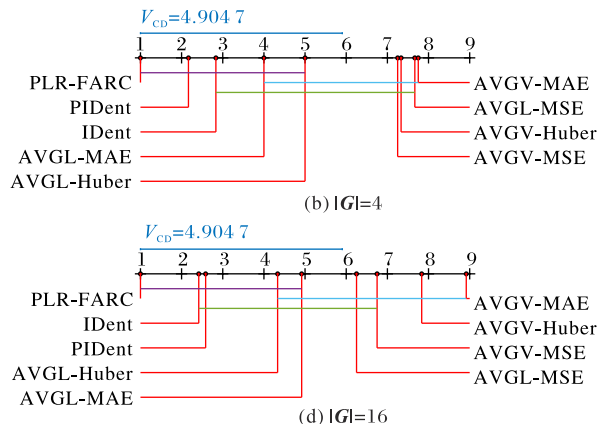


图 4 Nemenyi 检验结果

Fig. 4 Nemenyi test results

4 结语

偏标签回归是偏标签学习的延伸和拓展。不同于仅关注回归损失设计的偏标签回归算法, 本文算法从特征的角度

出发, 重点研究特征的特性差异及其在传递过程中的完整性。引入注意力模块以自动计算特征对模型训练的贡献度, 借助残差连接以降低特征信息在传递过程中的损失。在 6 个基准数据集上进行对比实验, 本文算法的性能优于对比算

法。实验结果表明,本文算法可行且有效。然而,本文基于均匀分布随机生成候选标签,未充分考虑标签与实例间的关联性,如何设计出更具合理性的候选标签是未来进一步研究的重点方向。

参考文献(References)

- [1] TIAN Y, YU X, FU S. Partial label learning: taxonomy, analysis and outlook[J]. *Neural Networks*, 2023, 161: 708-734.
- [2] JIA Y, PENG X, WANG R, et al. Long-tailed partial label learning by head classifier and tail classifier cooperation [C]// *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2024: 12857-12865.
- [3] LV J, XU M, FENG L, et al. Progressive identification of true labels for partial-label learning [C]// *Proceedings of the 37th International Conference on Machine Learning*. New York: JMLR. org, 2020: 6500-6510.
- [4] WANG D B, ZHANG M L, LI L. Adaptive graph guided disambiguation for partial label learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(12): 8796-8811.
- [5] LIAO Z, XIE Y, HU S, et al. Learning from ambiguous labels for lung nodule malignancy prediction [J]. *IEEE Transactions on Medical Imaging*, 2022, 41(7): 1874-1884.
- [6] XU Y Y, SHEN Y, WEI X S, et al. Weakly-supervised fine-grained recognition with partial label learning[C]// *Proceedings of the 31st International Joint Conference on Artificial Intelligence*. California: ijcai. org, 2022: 1502-1508.
- [7] YU X R, WANG D B, ZHANG M L. Partial label learning with emerging new labels[J]. *Machine Learning*, 2024, 113(4): 1549-1565.
- [8] GONG X, YANG J, YUAN D, et al. Generalized large margin k NN for partial label learning [J]. *IEEE Transactions on Multimedia*, 2022, 24: 1055-1066.
- [9] GU Y, CHEN Z, QIN Y, et al. DEER: distribution divergence-based graph contrast for partial label learning on graphs[J]. *IEEE Transactions on Multimedia*, 2024(Early Access): 1-16.
- [10] LIN G Y, XIAO Z Y, LIU J T, et al. Feature space and label space selection based on error-correcting output codes for partial label learning[J]. *Information Sciences*, 2022, 589: 341-359.
- [11] LIU L P, DIETTERICH T G. Learnability of the superset label learning problem[C]// *Proceedings of the 31st International Conference on Machine Learning*. New York: JMLR. org, 2014: 1629-1637.
- [12] LV J, LIU B, FENG L, et al. On the robustness of average losses for partial-label learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(5): 2569-2583.
- [13] XU N, LV J, GENG X. Partial label learning via label enhancement [C]// *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2019: 5557-5564.
- [14] CHENG X, WANG D B, FENG L, et al. Partial-label regression [C]// *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2023: 7140-7147.
- [15] YUAN X, HUANG L, YE L, et al. Quality prediction modeling for industrial processes using multiscale attention-based convolutional neural network [J]. *IEEE Transactions on Cybernetics*, 2024, 54(5): 2696-2707.
- [16] HUANG W, DENG Y, HUI S, et al. Sparse self-attention transformer for image inpainting[J]. *Pattern Recognition*, 2024, 145: No. 109897.
- [17] WANG D, ZHAI L, FANG J, et al. psoResNet: an improved PSO-based residual network search algorithm [J]. *Neural Networks*, 2024, 172: No. 106104.
- [18] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [19] WANG S, SHI J, YANG W, et al. High and low frequency wind power prediction based on Transformer and BiGRU-attention[J]. *Energy*, 2024, 288: No. 129753.
- [20] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers). Stroudsburg: ACL, 2019: 4171-4186.
- [21] LIU T, HONG J, WANG J, et al. Uniform distribution of zinc ions achieved by functional supramolecules for stable zinc metal anode with long cycling lifespan [J]. *Energy Storage Materials*, 2022, 45: 1074-1083.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc. , 2017: 6000-6010.
- [23] CHENG Y, QIAN K, MIN F. Global and local attention-based multi-label learning with missing labels[J]. *Information Sciences*, 2022, 594: 20-42.
- [24] YU X, ZHANG D, ZHU T, et al. Novel hybrid multi-head self-attention and multifractal algorithm for non-stationary time series prediction[J]. *Information Sciences*, 2022, 613: 541-555.
- [25] SHAFIQ M, GU Z. Deep residual learning for image recognition: a survey[J]. *Applied Sciences*, 2022, 12(18): No. 8972.
- [26] FERNÁNDEZ A J. Planning reliability demonstration tests with limited expected risks[J]. *Computers and Industrial Engineering*, 2022, 165: No. 107918.
- [27] NAIR P, VAKHARIA V, SHAH M, et al. AI-driven digital twin model for reliable lithium-ion battery discharge capacity predictions [J]. *International Journal of Intelligent Systems*, 2024, 2024: No. 8185044.
- [28] KARUNASINGHA D S K. Root mean square error or mean absolute error? use their ratio as well[J]. *Information Sciences*, 2022, 585: 609-629.
- [29] GE W, WANG Y, XU Y, et al. Causality-driven intra-class non-equilibrium label-specific features learning[J]. *Neural Processing Letters*, 2024, 56(2): No. 120.
- [30] DEMŠAR J. Statistical comparisons of classifiers over multiple data sets [J]. *Journal of Machine Learning Research*, 2006, 7: 1-30.

This work is partially supported by Anhui Provincial Natural Science Foundation (2108085MF216); Natural Science Research Project of Anhui Colleges and Universities (2024AH040175, 2024AH051099).

WU Haifeng, born in 1982, M. S., professor. His research interests include machine learning, data mining.

TAO Liqing, born in 1999, M. S. candidate. Her research interests include partial label learning, data mining.

CHENG Yusheng, born in 1969, Ph.D., professor. His research interests include rough set, semi-supervised learning, machine learning, data mining.