

# 公平性增强的决策树算法

姜文慧 叶剑虹 高灵婷 黄一凡

华侨大学计算机科学与技术学院 福建 厦门 361021

(jwh1010@qq.com)

**摘要** 在机器学习领域,模型的内在偏见问题日益受到关注,这些偏见往往源自训练数据的不平衡性或算法设计缺陷,从而导致某些群体在预测结果上受到不公正对待。为了解决这一问题,提出了一种公平性增强的决策树算法,通过引入公平性预处理方法,有效减少数据中的不平衡性,并且改变传统的决策树分裂标准,在决策树的分裂标准中综合考虑了分类准确性和公平性。所提方法旨在实现不同群体间预测结果的公平分配,减少模型决策中的偏见,确保所有个体得到公正对待。实验结果表明,所提出的方法在多种公平性度量标准下展现出良好的性能,显著降低了不同群体间的预测偏差,具有比现有传统算法更强的公平性纠偏性能。

**关键词** 机器学习;分类;决策树;公平性;预处理

**中图分类号** TP391

## Fairness-enhancing Decision Tree Algorithm

JIANG Wenhui, YE Jianhong, GAO Lingting and HUANG Yifan

College of Computer Science and Technology, Huaqiao University, Xiamen, Fujian 361021, China

**Abstract** In the field of machine learning, the problem of intrinsic biases in models has received increasing attention, and these biases often originate from imbalances in the training data or flaws in the algorithm design, which lead to unfair treatment of certain groups in the prediction results. To address this problem, this paper proposes a fairness-enhanced decision tree algorithm, which effectively reduces the imbalance in the data by introducing a fairness preprocessing method, and changes the traditional decision tree splitting criterion by integrating classification accuracy and fairness in the splitting criterion of the decision tree. The proposed method aims to achieve the fair distribution of prediction results among different groups, reduce the bias in model decision-making, and ensure that all individuals are treated fairly. Experimental results show that the proposed method demonstrates good performance under multiple fairness metrics, significantly reduces the prediction bias among different groups, and exhibits stronger fairness bias-correction performance than the existing traditional algorithms.

**Keywords** Machine learning, Classification, Decision tree, Fairness, Preprocessing

## 1 引言

现实世界中存在一些不公平现象,比如在银行审批贷款的过程中,算法的偏差会导致基于申请人的性别、种族等因素做出有偏见的决策,使得某些群体受到不公平对待。如表1所列,普通决策树将性别作为主要分裂点,基于性别而非收入状况来决定贷款资格,从而预测一位30岁、年收入60000美元的女性不具备贷款资格,而事实上贷款资格与收入状况更相关。

表1 一个简单的例子(局部)

Table 1 A simple example(partial)

年龄	年收入	性别	是否有贷款资格
25	\$ 50 000	男	是
30	\$ 60 000	女	否

在处理机器学习中的歧视问题时,研究者们已经探索了多种策略来减少或消除对某些群体的不公平待遇<sup>[1]</sup>。这些不公平现象主要由数据中固有的偏见和数据与算法之间错综复

杂的交互所引起<sup>[2]</sup>。为了应对这一挑战,学者们进行了广泛而深入的研究,旨在减少或消除算法偏见,提升模型的公平性<sup>[3]</sup>。这些研究通常采用以下3种策略:数据预处理、过程中处理以及后处理<sup>[4]</sup>。

数据预处理策略,是指在训练模型之前对数据进行处理,以减少数据中的偏见<sup>[5]</sup>。具体而言,数据预处理策略包括多种方法,例如重新加权,为不同的群体分配不同的权重,以减轻对群体的偏见<sup>[6]</sup>。分配权重的常见方式是基于实例的敏感属性值和类别的预期和观察概率,如果观察概率低于预期概率,则会分配更高的权重,以中和歧视<sup>[7]</sup>。又例如重新采样,它能平衡数据集<sup>[8]</sup>,可以采用子采样(从多数群体中移除一些实例)或过采样(复制少数群体的实例)来确保数据集中各个敏感属性子群体的公平代表性。这两种方法都可以调整数据分布,使得模型学习时不会过度偏向某一特定群体。Friedler等<sup>[9]</sup>的比较研究表明,预处理技术如重新加权和过采样能有效提升模型的公平性。数据清洗也是一种常用的预处理策

基金项目:福建省科技厅引导性项目(2024H0014(2024H01010100))

This work was supported by the Science and Technology Planning Project of Fujian Province, China(2024H0014(2024H01010100)).

通信作者:叶剑虹(leafever@163.com)

略,它能识别并移除或修正那些导致歧视的数据点<sup>[10]</sup>。除此之外,González-Zelaya 等提出的 FAIR-MDAV 方法<sup>[11]</sup>是一种专门设计用于同时实现数据隐私保护和公平性的数据预处理策略,它通过微聚合(Microaggregation)技术将数据分组,并在这些分组内部进行公平性纠正。实验结果显示,FAIR-MDAV 能够有效降低模型在不同群体上的公平性差异。数据预处理策略有关方法的理论依据是:如果分类器是在无歧视数据集上训练的,它将学习到更加公平的模式,从而在预测时不会延续原有数据集的不公平现象。

过程中处理策略通过直接在算法中引入公平性约束,以确保学习过程中的公平性<sup>[12]</sup>。这种策略包括公平感知学习<sup>[13]</sup>,即在模型的训练过程中引入公平性指标如机会平等。Hardt 等提出了机会平等的概念,这是一种在模型训练过程中考虑公平性的指标<sup>[14]</sup>。此外,正则化技术也被用于促进公平性,通过添加正则化项来惩罚模型对某些群体的不公平预测<sup>[15]</sup>。Zafar 等提出了一种正则化方法,以减少分类算法在不同群体上产生的不公平差异<sup>[16]</sup>。

另一些研究者选择采取后处理策略,即在模型训练完成后,对模型的预测结果进行调整,以提高公平性。例如结果校正,调整模型的预测结果,以确保不同群体的预测结果更加公平<sup>[17]</sup>。这种方法涉及对模型的预测结果进行调整,以确保不同群体的预测结果更加公平。

研究者 Kamiran 等提出了一种新的技术,称为“歧视意识决策树学习”(Discrimination Aware Decision Tree Learning)<sup>[18]</sup>,旨在从存在偏见的历史数据中训练出一个既具有预测准确性,又不基于敏感属性进行歧视的分类器。这种方法通过改变决策树学习过程中的分裂标准和引入了一种称为“重标记”(Leaf Relabeling)的后处理技术,进一步将非歧视约束深入集成到决策树学习器中,重标记技术通过调整决策树叶节点的类别标签,以降低歧视度,同时尽量最小化对预测准确性的影响。但是 Kamiran 等的方法还存在不足之处,主要体现在数据层面和算法层面。在数据层面,该方法的效果高度依赖于训练数据的质量和分布,如果训练数据本身存在严重的歧视或偏见,模型将难以学习到公平的分类器,因为算法会学习并复制这些固有偏见,从而影响模型的公平性。在算法层面,重标记技术在降低歧视度的过程中,容易矫枉过正,对另一方造成歧视。

为了解决上述问题,本文提出了一种公平性增强的决策树算法。一方面,首先采用 FAIR-MDAV 方法在决策树开始训练前,对训练数据集进行公平性预处理,将数据分成多个小集合,在小集合内进行公平性纠正,避免了 Kamiran 等方法的效果高度依赖于训练数据的质量和分布的问题。另一方面,通过改进传统决策树分裂标准,将衡量机器学习模型公平性的指标 Demographic Parity 引入决策树分裂标准中,使改进后的决策树标准同时综合分类的准确性和公平性,并引入调节参数,使模型的决策灵活性增强,以适应不同应用场景。本文方法在公开数据集上与其他方法进行对比,结果表明,本文方法能有效降低不公平性,并且不损失过多的预测性能,获得更公平的分类效果。通过减少决策过程中的歧视,本文方法有助于创建一个更加多样和包容的环境,为社会的整体进步和创新做出积极贡献。

## 2 理论知识

本文考虑的场景是一个具有标签的数据集,其中存在因敏感属性导致的内在预测偏差。本文的目标是从这一有偏数据集中构建一个公平的预测模型,使得模型在未来的预测中不易受到敏感属性的影响,从而减少潜在的不公正现象。在机器学习领域,敏感属性(Sensitive Attribute,简称 SA)是指一些与个人身份相关的特征,如种族、性别、年龄等,这些特征在法律或社会伦理上被认为不应成为歧视的依据。在机器学习公平性讨论中,主要针对二元敏感属性进行研究,二元敏感属性的值可以分为两种情况:一种是“有利”(Favoured,简称 F),表示在决策中某些群体可能因其敏感属性而得到优待;另一种是“不利”(Unfavoured,简称 U),则指某些群体可能因敏感属性而遭受不公正待遇。例如在招聘场景中,定义敏感属性为性别,如果模型更倾向于选择男性候选人,那么女性候选人就属于不利群体。

在构建决策树的过程中,节点划分标准是关键因素<sup>[19]</sup>,节点划分的标准以及公平性相关评价指标如下,以下定义来源于文献<sup>[20]</sup>。

**定义 1(信息增益)** 信息增益用于衡量分裂对分类准确度的贡献:

$$IGC(D, A) = H_{\text{Class}}(D) - \sum_{i=1}^k \frac{|D_i|}{|D|} H_{\text{Class}}(D_i) \quad (1)$$

**定义 2(歧视度量)** 歧视度量是衡量分类器在特定敏感属性上的预测偏差:

$$DISC_{SA}(C, D) := \frac{|x \in D | x \cdot SA = F, C(x) = 1|}{|x \in D | x \cdot SA = F|} - \frac{|x \in D | x \cdot SA = U, C(x) = 1|}{|x \in D | x \cdot SA = U|} \quad (2)$$

其中, $C$ 表示分类器; $D$ 表示数据集; $x$ 表示数据集中单个实例; $SA$ 表示敏感属性。 $C(x)$ 表示分类器对实例  $x$  的预测结果, $C(x)=1$ 表示分类结果为正,歧视度量越低,代表分类器结果公平性越好。

**定义 3(敏感属性增益)** 敏感属性增益衡量分裂对敏感属性的影响:

$$IGS = H_{SA}(D) - \sum_{i=1}^k \frac{|D_i|}{|D|} H_{SA}(D_i) \quad (3)$$

其中, $D$ 是原始数据集, $H_{SA}(D)$ 表示敏感属性的熵, $k$ 是分裂产生的子集数量, $D_i$ 是由分裂标准产生的数据子集。低  $IGS$  值表明模型在分裂时,对于敏感属性值为有利群体和不利群体的分类结果较为相似。

统计均等性要求在预测分类任务中,模型对不同群体的正类预测概率相等。这意味着,不论个体的敏感属性(如种族、性别等)如何,模型对各个群体的预测结果应该是公平的。

**定义 4(统计均等性(Demographic Parity, DP))** 统计均等性又称人口统计公平性,数学表达式如下:

$$P(\hat{Y}=1 \mid SA=U) = P(\hat{Y}=1 \mid SA=F) \quad (4)$$

分类器的统计均等得分是通过计算在敏感属性值为不利群体和有利群体时,被分类为正面结果的概率之差的绝对值来确定的。得分越低,表示分类器在不同群体间的公平性越好。

**定义 5**(统计均等得分(Demographic Parity Score,DPS)) 统计均等性得分又称人口统计公平性得分,数学表述如下:

$$DPS(\hat{Y})=|P(\hat{Y}=1 \mid SA=U)-P(\hat{Y}=1 \mid SA=F)| \quad (5)$$

其中, $P(\hat{Y}=1 \mid SA=U)$ 表示在敏感属性值为不利群体时,被分类为正面结果即 $\hat{Y}=1$ 的概率; $P(\hat{Y}=1 \mid SA=F)$ 表示在敏感属性值为有利群体时,被分类为正面结果即 $\hat{Y}=1$ 的概率。

**定义 6**(概率均等得分(Equalized Odds Score,EOS)) 概率均等得分衡量了模型在不同敏感属性群体间预测性能的不一致性。分数越低,表示模型在不同群体间的公平性越高,预测结果越不依赖于敏感属性。具体来说,该分数考察在给定真实标签的情况下,不同敏感属性群体被正确分类的概率是否相等。数学表述如下:

$$EOS(\hat{Y})=\sum_{i \in \{0,1\}} |P(\hat{Y}=i \mid Y=i, SA=U)-P(\hat{Y}=i \mid Y=i, SA=F)| \quad (6)$$

### 3 算法研究

#### 3.1 预处理方法(FAIR-MDAV 方法)

传统的机器学习方法中,模型往往会放大数据中对某些群体的不公平偏见。如图 1 所示,这个决策树是由一个简化的成人收入数据集训练得到的,该数据集包含年龄、性别、教育程度等多种特征,目标是根据这些特征预测个体的年收入是否超过 50000 美元。最终得到的传统决策树反映了数据中的偏见模式,它直接使用性别(sex 属性)作为分裂标准,这种做法会对某一性别的个体产生不利的预测结果。在该决策树的叶子节点中,被预测为年收入超过 50000 美元的样本均为男性,忽略了女性同样具备获得高收入的能力。这种偏见导致了性别歧视,因为它基于性别对收入能力进行了不公平的划分。

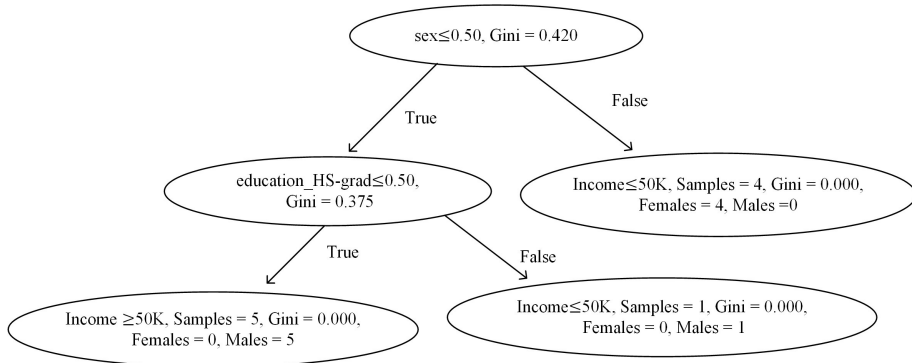


图 1 普通决策树的不公平现象

Fig. 1 Unfairness of common decision trees

这种偏见源于训练数据可能包含历史累积的不公平实践和歧视,模型在学习过程中会继承甚至强化这些偏见,导致对敏感属性(如性别、种族、年龄等)群体的歧视,对特定群体造成不利影响。因此,在机器学习应用中,引入数据公平预处理很有必要,以对训练数据进行消除或减轻不公平偏见。

本文的数据预处理方法采用 FAIR-MDAV 方法,其在数据预处理中同时实现了公平性和隐私保护。首先将要进行训练的数据集划分为若干个“公平小单元”(Fairlet),每个小单元内的样本在敏感属性上的分布被设计为接近整体数据集的分布,从而确保组内不同群体的平衡。然后,在这些小单元内部进行公平性纠正,可以选择两种方式:正向纠正(将不利群体的负例转换为正例)和负向纠正(将有利群体的正例转换为负例),从而在不显著影响数据集整体标签分布的情况下提升模型的公平性。其具体步骤如算法 1 所示。

#### 算法 1 FAIR-MDAV

输入:训练数据集  $D$ ,小单元内不利群体记录的数量  $m$  和有利群体记录的数量  $n$

输出:经过公平性处理后的训练集  $Data$

1.  $G \leftarrow \{\}$
2. while  $|\{x \in D \mid x \cdot SA = U\}| \geq m$  and  $|\{x \in D \mid x \cdot SA = F\}| \geq n$  do
3.  $x_{mean} \leftarrow \text{Mean}(\{x \in D\})$ ;
4.  $x_r \leftarrow \text{argmax}_{x \in D}(\text{distance}(x, x_{mean}))$ ;
5.  $g \leftarrow \text{Append}(x_r)$ ;

6. if  $x_r \cdot SA = F$  then
7.  $g_u \leftarrow \text{GetNearestU}(x_r, m)$ ;
8.  $g_f \leftarrow \text{GetNearestF}(x_r, n-1)$ ;
9. else
10.  $g_u \leftarrow \text{GetNearestU}(x_r, m-1)$
11.  $g_f \leftarrow \text{GetNearestF}(x_r, n)$ ;
12. end if
13.  $g \leftarrow \text{Append}(g_u, g_f)$ ;
14.  $D \leftarrow \text{Drop}(D, g)$ ;
15.  $G \leftarrow \text{Append}(G, g)$ ;
16. end while
17. for  $g$  in  $G$  do
18.  $fpr \leftarrow \text{PositiveRatio}(g, 1)$ ;
19.  $upr \leftarrow \text{PositiveRatio}(g, 0)$ ;
20. List  $U_0 \leftarrow \{x \text{ in } g \mid x \cdot SA = U \text{ and } y(x) = 0\}$ ;
21. List  $F_1 \leftarrow \{x \text{ in } g \mid x \cdot SA = F \text{ and } y(x) = 1\}$ ;
22. if  $nc$  is False then
23. while  $upr < \tau \cdot fpr$  and  $|List U_0| > 0$  do
24.  $y(x) = 1$  for  $x$  in List  $U_0$
25. else
26. while  $upr < \tau \cdot fpr$  and  $|List F_1| > 0$  do
27.  $y(x) = 0$  for  $x$  in List  $F_1$
28. end if
29. end for

30. Data ← Concatenate({g in G});

31. return Data

第1行初始化了一个集合G,后续用来存储由数据集D划分成的多个小单元g。

第2—16行将数据集D划分为由多个小单元g组成的集合G。每个小单元g包含m个不利群体记录和n个有利群体记录,使得 $\frac{m}{n}$ 的比例尽可能接近数据集D中不利群体记录和有利群体记录的比例,同时满足 $m+n=k$ ,其中k为聚类大小(即小单元大小),以确保每个小单元在一定程度上反映了整体数据集的分布特征。第3行计算数据集D的平均记录,记为 $x_{mean}$ 。第4—5行找到数据集中距离平均记录最近的记录 $x_r$ 。第6—8行根据 $x_r$ 的敏感属性(SA)的值,定义辅助集合 $g_f$ 和 $g_u$ ,如果 $x_r$ 属于有利群体,则 $g_f$ 包含距离 $x_r$ 最近的n-1个SA取值为有利的记录,而 $g_u$ 包含距离 $x_r$ 最近的m个SA取值为不利的记录。第10—11行如果 $x_r$ 属于不利群体,则 $g_f$ 包含距离 $x_r$ 最近的n个SA取值为有利的记录,而 $g_u$ 包含距离 $x_r$ 最近的m-1个SA取值为不利的记录。第13行由 $x_r \cup g_u \cup g_f$ 组成一个新的小单元g,这个小单元包含 $x_r$ 的最近邻,并且保持了数据集D中的SA分布比例。第14—15行将新形成的小单元g添加到G中,即数据集D的小单元划分集合。然后从D中移除这些记录,并在剩余的记录上重复上述划分过程,直到无法形成更多的小单元,或者剩余记录不足以形成一个完整的小单元。

第17—29行在每个小单元内进行局部校正,局部校正的目的是调整每个小单元中的数据,以减少不同SA值之间的预测偏差和不公平性,校正过程是利用while循环迭代进行的。其中第18—19行计算敏感属性为不利群体和有利群体的记录在这个小单元中被分类为正例的比例(PositiveRatio,简称PR),分别是 $upr$ 和 $fpr$ 。第20—21行将该小单元内记录根据敏感属性和类标签分类。第22—28行设置调节参数 $\tau$ 控制校正的程度,如果计算后发现 $upr < \tau \cdot fpr$ ,有两种校正方法可供选择,一种是正向校正,将一些属于不利群体的被分类为负例的记录(标签为0的记录)重新标记为正例(标签改为1),以提高不利群体被分类为正例的比例。 $nc$ 为False代表正向校正。另一种是负向校正,将一些属于有利群体的被分类为正例的记录(标签为1的记录)重新标记为负例(标签改为0),以减少两群体之间正例比例的差异。 $nc$ 为True代表负向校正。第30—31行将经过校正后的所有小单元放入集合G,由集合G内数据构成经过校正的数据集Data。

算法1(FAIR-MDAV方法)的总体时间复杂度为 $O(N^2 p/k)$ ,其中N表示数据集D中的记录总数,p表示每条记录的属性数量(即维度数),k表示聚类大小(每个小单元包含的记录数量)。因为构建小单元的过程需要多次计算记录之间的距离,每次计算的复杂度为 $O(Np)$ ,且该过程需要重复约 $N/k$ 次(因为每次迭代处理k条记录,直到处理完所有N条记录),所以累计的时间复杂度为 $O(N^2 p/k)$ 。其他步骤的复杂度为常数级别和线性级别,对总体时间复杂度的影响较小。尽管时间复杂度理论上存在与N的二次关系,但在实际应用中,只要合理选择k并采用策略,比如使用高效的数据结构、降维技术或并行处理来加速,算法运行效率仍然很高。因此,FAIR-MDAV方法在确保数据公平性的同时,能够有效地处理大型数据集。

例1 表2为算法1生成的一个小单元。对于表2中该小单元的处理,FAIR-MDAV方法首先确定敏感属性“性别”中不同群体的分布比例。该单元中男性占2条记录,女性占1条记录,且男性的“高收入”比例为100%,而女性为0%,存在显著的比例差异。假设调节系数为1,意味着在小单元内女性群体的“高收入”比例需要与男性保持一致,为此,算法进行正向校正,将女性的“否”记录修改为“是”,从而实现了小单元在性别属性上的公平性。通过对每个小单元内的记录进行类似校正,FAIR-MDAV方法逐步平衡了数据集中不同敏感属性群体的比例,使得在整体数据集上不同群体的预测结果趋于一致,从而在全局范围内实现了公平性的目标。

表2 FAIR-MDAV算法的例子

Table 2 Example of FAIR-MDAV algorithm

年龄	性别	...	判定收入是否为高收入
41	男	...	是
30	男	...	是
31	女	...	否

### 3.2 促进公平性的决策树分裂标准

原始的决策树分裂标准(如信息增益等)在选择最优分裂点时仅关注纯度的提升和分类性能的最大化,而忽视了模型在敏感属性上的公平性问题。具体来说,首先,这些分裂标准没有考虑敏感属性(如性别、种族)对决策结果的影响,在决策过程中易引入隐性的偏见,使得模型更依赖某些敏感属性来区分群体,从而在不同敏感属性的群体间产生预测不平衡性,如图1所示的普通决策树不公平现象。其次,即使数据集已经过预处理,敏感属性的直接依赖性或不均问题有所缓解,但是原始分裂标准在模型训练时仍会引入潜在偏差,倾向于选择与敏感属性相关的特征,从而导致有利群体和不利群体之间的预测结果不平衡。例如,在招聘场景中,即使预处理方法已经平衡了男性和女性在“是否有管理经验”上的比例,但原始决策树仍容易偏向选择与“性别”存在隐性关联的其他特征(如“工作年限”或“是否全职”),因为传统分裂标准往往更关注纯度提升而忽视潜在的群体差异。因此,仅仅依靠数据预处理难以完全消除这些隐性偏差,仍需要在模型训练时引入促进公平性的分裂标准,从根本上约束模型对某些特征的偏向性选择,确保模型在提升分类性能的同时,保持对不同群体的公平性。值得注意的是,尽管Kamiran等提出了结合信息增益和敏感属性增益结合(IGC+IGS)的分裂标准,但该分裂标准无法单独使用,必须与后续的叶子节点重标记技术配合才能实现公平性。

为了弥补这些缺陷,本文引入一个综合考虑公平性和分类准确性的分裂标准,该标准结合了传统的信息增益(定义1)和人口统计公平性(定义4)变化量,以在提升模型分类纯度的同时,改善模型的公平性。通过这种方法,可以确保敏感属性群体在左右子节点中的分布保持公平,从而实现更加全面和公正的决策过程。

在每次决策树的潜在的分裂中,需要计算分裂产生的信息增益 $IGC(D,A)$ 和人口统计公平性差异 $\Delta DPS$ 。

对于当前节点,计算分裂产生的信息增益,如式(1)所示,以及计算分裂前后的人口统计公平性差异。首先定义敏感属性SA有两个取值,不利群体表示为 $SA=U$ ,有利群体表示为 $SA=F$ 。

在每次潜在的分裂中,比较在左右子节点内这两个群体的正类预测率差异,以计算人口统计公平性差异。由于在决策树分裂过程中无法直接得知分裂节点中各群体的最终正类预测率,故采用每个节点中不同群体的正类样本比例来近似正类预测率。这种方法的逻辑在于,虽然无法直接得到节点中的预测值,但可以通过样本标签的比例来估计预测的潜在偏差。

分裂前,节点中的人口统计公平性分数  $DPS_{\text{before}}$  可以表示为:

$$DPS_{\text{before}} = |P(\hat{Y}=1|SA=U) - P(\hat{Y}=1|SA=F)| \approx \left| \frac{n_U^+}{n_U} - \frac{n_F^+}{n_F} \right| \quad (7)$$

其中,  $n_U^+$  和  $n_U$  分别表示当前节点中不利群体的正类样本数量和不利群体总样本数量;  $n_F^+$  和  $n_F$  分别表示当前节点中有利群体的正类样本数量和有利群体总样本数量。

由于本文处理的是二分类任务,计算左右子节点中敏感属性群体之间的正类预测率之差与计算负类预测率之间的差异具有相似的意义,关注正类和负类的预测率差异,均可有效验证模型在不同群体之间的公平性。

例2 设节点内总样本数为150个,包括正类样本115个,负类样本35个。将这150个样本分为两个群体:不利群体  $U$  和有利群体  $F$ , 其中不利群体  $U$  有100个样本,包括70个正类样本和30个负类样本,正类比例为0.70;有利群体  $F$  有50个样本,包括45个正类样本和5个负类样本,正类比例为0.90。因此,两个群体之间正类比例差异为0.20。同样地,经计算得到不利群体  $U$  的负类比例为0.30,有利群体  $F$  的负类比例为0.10,两个群体之间负类比例差异也为0.20。这表明,在二分类问题中,如果两个群体的正类比例存在0.20的差异,那么负类比例之间也会存在相同数值的差异,且这两个差异是完全对称的,可以通过分析任一类别的比例差异来推断另一类别的差异,从而全面评估模型在不同群体间的公平性。

接下来,计算左右子节点的人口统计公平性分数。

为了预防分裂后的子节点中出现不利群体  $U$  和有利群体  $F$  中某一群体的样本数量为零的情况,在分裂过程中会筛选出那些左右子节点中均包含不利群体  $U$  和有利群体  $F$  样本的分裂点。如果某个潜在的分裂点会导致分裂后左子节点或右子节点中某一群体的样本数量为零,则该分裂点将被排除。这种策略确保了在每次分裂后,左右子节点中敏感属性群体的分布保持公平,避免因样本不足而导致的人口统计公平性差异无法计算或失真的问题。

左子节点的人口统计公平性分数  $DPS_{\text{left}}$  的取值范围为0~1,由计算不利群体和有利群体的正类预测率之差的绝对值得到:

$$DPS_{\text{left}} = |P(\hat{Y}_{\text{left}}=1|SA=U) - P(\hat{Y}_{\text{left}}=1|SA=F)| \quad (8)$$

同样的方法应用于右子节点,右子节点的  $DPS_{\text{right}}$  的取值范围也为0~1,由计算节点内不利群体和有利群体的正类预测率之差的绝对值:

$$DPS_{\text{right}} = |P(\hat{Y}_{\text{right}}=1|SA=U) - P(\hat{Y}_{\text{right}}=1|SA=F)| \quad (9)$$

分裂后的加权平均  $DPS_{\text{after}}$  为:

$$DPS_{\text{after}} = \frac{n_{\text{left}} \times DPS_{\text{left}} + n_{\text{right}} \times DPS_{\text{right}}}{n_{\text{total}}} \quad (10)$$

其中,  $n_{\text{left}}$  为分裂后左子节点的样本数量,  $n_{\text{right}}$  为分裂后右子节点的样本数量,  $n_{\text{total}}$  为2个子节点的样本数量总和。

再由  $DPS_{\text{after}}$  和  $DPS_{\text{before}}$  得到分裂前后的人口统计公平性变化  $\Delta DPS$ 。

在评估所有潜在的分裂属性时,将信息增益和分裂前后的人口统计公平性变化  $\Delta DPS$  结合起来,形成一个综合的分裂标准。为了在分裂标准中平衡分类准确性和公平性,本文引入了一个调节参数  $\lambda$  ( $0 \leq \lambda \leq 1$ ),最终分裂增益定义为:

$$Gain = (1-\lambda) \cdot IGC(D,A) - \lambda \cdot \Delta DPS \quad (11)$$

通过调整  $\lambda$  的值,模型可以根据具体应用场景的需求,优先考虑分类准确性或公平性。在  $\lambda=0$  时,模型将完全专注于信息增益,以达到最高的分类性能;而在  $\lambda=1$  时,模型则优先考虑公平性,以最小化群体间的预测差异。模型所要达到的目标为:

$$\begin{cases} \text{Maximize } IGC(D,A), & \text{when } \lambda=0 \\ \text{Minimize } \Delta DPS, & \text{when } \lambda=1 \end{cases} \quad (12)$$

这种灵活的调节机制为决策树的构建提供了更为全面的视角,使模型在面对多样化的数据环境时,能够实现公平与准确有效平衡,确保模型不仅关注准确性,还重视不同群体之间的公平性,为构建公正、准确的机器学习模型奠定了基础。

### 3.3 公平性增强的决策树构造

原始的训练数据集经过 FAIR-MDAV 方法预处理后,得到公平性增强的数据集,再使用促进公平性的分裂标准用于决策树的构造,即构成了本文所提出的公平性增强的决策树算法,其详细步骤如算法2所示。

#### 算法2 Fairness-Enhancing Decision Tree Construction

输入: 经过处理的训练数据集 Data, 决策树的根节点 node, 调节参数  $\lambda$ , 敏感属性 SA

输出: 公平性增强的决策树模型

1. function TreeConstruction(node, Data,  $\lambda$ , SA):
2. if IsPure(Data) or maximum depth reached then
3. return // 停止构造的条件
4. end if
5. best\_gain =  $-\infty$
6. best\_split = None
7. for each attribute A in Data do
8. split\_points = GetSplitPoints(Data, A)
9. for each split\_point in split\_points do
10. split Data into left\_data, right\_data
11. calculate IGC
12. calculate DpChange
13. gain =  $(1-\lambda) * IGC - \lambda * DpsChange$
14. if gain > best\_gain then
15. best\_gain = gain
16. best\_split = (A, left\_data, right\_data)
17. create the node
18. end if
19. end for
20. end for
21. TreeConstruction(node, left\_child, best\_split[1],  $\lambda$ , SA)
22. TreeConstruction(node, right\_child, best\_split[1],  $\lambda$ , SA)

算法2第1行对输入数据进行处理,包括训练数据集 Data、敏感属性 SA 以及调节参数  $\lambda$ 。第2-4行在递归分裂过程中,检查当前数据集是否纯净(即所有样本是否属于同一

类别)或是否达到最大深度,如果满足条件,则停止构造。第5-6行初始化最佳分裂增益  $best\_gain$  和最佳分裂点  $best\_split$ 。第7-10行遍历数据集中的每个属性,获取该属性的所有可能分裂点,对于每一个分裂点,将数据集划分为左子集和右子集。第11-12行计算当前分裂点的信息增益和人口统计公平性变化。第13行结合信息增益和人口统计公平性变化,计算综合增益。第14-18行如果当前综合增益优于最佳分裂增益,则更新最佳分裂增益  $best\_gain$  和最佳分裂点  $best\_split$ ,并创建节点。第21-22行对左子集和右子集递归调用  $TreeConstruction$  函数,继续构建决策树。

算法2的主要时间复杂度来源于属性分裂点选择及计算分裂产生的信息增益和人口统计公平性差异,假设数据集包含  $N$  条记录,每条记录有  $p$  个属性。对于决策树的每次分裂操作,算法需要在每个属性上找到可能的分裂点,并评估分裂产生的人口统计公平性差异和信息增益,分裂点数量取决于属性的类型,对于连续属性,最坏情况下分裂点的数量可能接近于  $N-1$ ,每个不同的值都可以是一个潜在的分裂点,对于离散属性,分裂点的数量等于属性的不同值的数量,这通常远小于  $N$ 。因此,每个属性最多有  $N$  个可能的分裂点,而对于每个分裂点的评估包括信息增益和公平性变化的计算,其时间复杂度为  $O(N)$ ,因此在每个属性上的分裂评估总时间为  $O(N^2)$ 。由于数据集有  $p$  个属性,所以在当前节点上找到最优分裂点的总体时间复杂度为  $O(p \times N^2)$ 。算法通过递归方式构建决策树,每次递归分裂会将数据集分为左子集和右子集,直到达到树的最大深度或所有节点的纯度满足停止条件。设树的最大深度为  $D$ ,因此,最坏情况下,公平性增强决策树构建算法的总体时间复杂度为  $O(p \times N^2 \times D)$ 。然而,需要注意的是,随着树的生长,每次分裂后左右子集的数据量都会减少,因此实际上并不总是需要处理全部  $N$  个样本,实际的时间复杂度会低于这个理论上的上限。

## 4 实验结果及分析

本文的实验环境为 Intel Core i7-10700 CPU,时钟速度 2.9 GHz,16 GB 机带 RAM,软件环境为 windows 11 操作系统,程序使用 Python 语言实现。

### 4.1 数据集

实验选取了 UCI 数据库中的 Statlog,Adult 和 Bank 数据集作为实验数据。Statlog 数据集是为了解决银行信用卡评估问题而收集的,包含了有关客户信用的各种信息,用于预测客户是否能够按时还款。Adult 数据集是一个公开可用的美国人口普查样本数据,通常用于分类任务。该数据集中的每个样本包含多种特征,如性别、工作类别、教育水平等,主要用于预测个体的年收入是否超过 50 000 美元。Bank 数据集源自葡萄牙某银行的营销活动,它包含了与客户相关的多种特征,如年龄、婚姻状态等,主要用于预测客户是否会接受定期存款产品。这 3 个数据集在公平性研究上具有重要意义,在相关研究中被广泛应用,以探索和修正算法中的潜在偏见,数据集的相关信息如表 3 所列。

对于 Statlog 数据集相关实验,本文选取性别作为敏感属性,性别  $\in \{\text{男性}, \text{女性}\}$ ,设男性为有利群体  $F$ ,女性设定为对照组  $U$ 。对于 Bank 数据集相关实验,本文选取年龄作为敏感属性,将年龄属性离散化,设 25-60 岁为有利群体  $F$ ,其他年

龄段为对照组  $U$ 。对于 Adult 数据集相关实验,本文选取性别作为敏感属性,性别  $\in \{\text{男性}, \text{女性}\}$ ,设男性为有利群体  $F$ ,女性设定为对照组  $U$ 。

表 3 数据集信息

Table 3 Datasets information

数据集名称	样本数	特征数
Statlog	1 000	20
Adult	30 913	10
Bank	45 203	14

### 4.2 实验相关设置

实验对类别型特征进行独热编码,用 FAIR-MDAV 进行预处理,其中校正系数  $\tau$  设为 1,每个小单元  $g$  包含 1 个不利群体记录和 2 个有利群体记录。决策树最大深度设置为 3。为了提高结果的可信度,本文实验采用五折交叉验证方法,并对所得性能指标的平均值进行分析与对比。

### 4.3 评价指标

采用准确率(Accuracy)作为分类性能评价指标,表示正确预测的样本数占总样本数的比例,它反映的是模型预测的正确性。其表达式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

其中,  $TP$ (True Positives)表示真正例,即模型正确预测为正类的样本数;  $TN$ (True Negatives)表示真负例,即模型正确预测为负类的样本数;  $FP$ (False Positives)表示假正例,即模型错误预测为正类的样本数;  $FN$ (False Negatives)表示假负例,即模型错误预测为负类的样本数。

采用歧视度量(DISC,见式(2))作为公平性评价指标,统计均等得分(DPS,见式(5)),概率均等得分(EOS,见式(6))。

### 4.4 实验结果与分析

#### 4.4.1 本文方法在不同调节参数下的表现

为了验证本文提出的综合分裂标准在不同调节参数  $\lambda$  下的表现,并最终找到最优参数,在两个数据集上评估模型在分类准确性和公平性方面的性能变化。 $\lambda$  从 0.1 到 0.9 的范围内,以 0.1 为步长调整调节参数,共计 9 个不同的  $\lambda$  值。对于每个  $\lambda$  值,每次实验训练模型并记录其在相关评价指标下的表现。

首先选择在小规模的 Statlog 数据集上进行实验,如图 2 所示,可以看出:随着调节参数的增加,公平性评价指标 EOS 和 DPS 呈下降趋势,DISC 越来越趋近于 0,表明模型在不同群体间的预测更加公平;分类准确率略有下降,但下降幅度较小,表明在实现公平性的过程中,未牺牲过多的预测性能,在  $\lambda=0.8$  时,模型实现了分类性能与公平性的良好平衡。

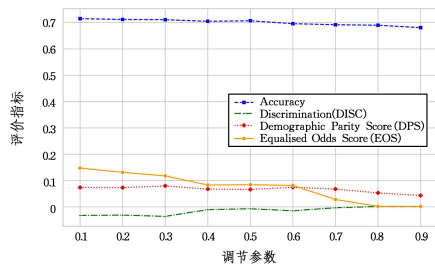


图 2 Statlog 数据集上不同调节参数下模型的表现

Fig. 2 Model performance under different tuning parameters on Statlog dataset

为了进一步评估模型在更复杂、实际场景中的表现,将实

验扩展到数据规模更大的 Adult 数据集和 Bank 数据集。

模型在 Adult 数据集上的表现如图 3 所示,可以看出随着  $\lambda$  的增大,不同指标的表现有显著变化。对于公平性评价指标,随着  $\lambda$  的增加,评价指标 DPS 逐渐下降并趋于稳定,表明随着公平性约束的引入,模型在不同群体之间的预测更加接近,公平性得到了改善。在  $\lambda$  取值较小时,DISC 和 EOS 这两个指标较高,表明模型存在明显的群体偏差,随着  $\lambda$  的增加,DISC 和 EOS 总体呈下降趋势,在  $\lambda > 0.5$  之后,DISC 越来越接近 0,说明模型公平性得到改善。通常,这种公平性的改善往往伴随着分类性能的下降,从图中可知,随着调节参数  $\lambda$  增大,分类准确率有所下降,但是下降幅度并不大,这反映了本文方法在公平性和准确性之间的良好平衡。当  $\lambda = 0.6$  时,模型在多个指标间达到了较好的平衡点,此时的分类准确率保持在较高水平,但公平性指标均显著降低,公平性评价指标 DISC 为  $-0.025$ ,DPS 为  $0.021$ ,EOS 为  $0.041$ ,说明在该参数值下,模型能够在保证一定分类性能的同时,显著提高不同群体之间的公平性。

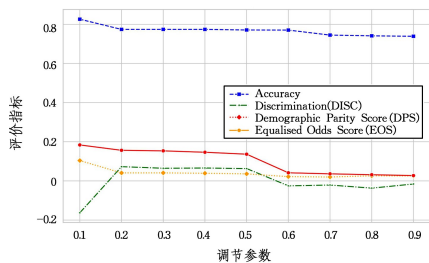


图 3 Adult 数据集上不同调节参数下模型的表现

Fig. 3 Model performance under different tuning parameters on Adult dataset

模型在 Bank 数据集上的表现如图 4 所示,在分类评价指标上,本文模型的表现与在 Adult 数据集上相似,随着  $\lambda$  的增大,准确率整体呈小幅下降的趋势,但其值在 0.8 附近。对于公平性评价指标,在  $\lambda$  取值为 0.1 时,模型在 Bank 数据集上的公平性指标 (DISC, DPS 和 EOS) 均处于较高水平,表明初始模型在群体间存在较大的偏差。但是在  $\lambda$  取值 0.2 后,模型的公平性指标已经有了显著改善,DISC, DPS 和 EOS 开始大幅下降,表明引入公平性约束降低了模型的群体偏差。随着  $\lambda$  的增大,DISC 和 DPS 总体呈稳定下降趋势,但当  $\lambda$  在 0.6~0.8 这个范围内时,评价指标 EOS 的值出现波动,在这个过程中,模型努力平衡公平性和分类性能,这种探索往往导致性能的短暂波动,直到找到一个相对稳定的状态。经过多指标的综合分析,当  $\lambda = 0.5$  时,模型分类性能较好,并且模型在不同群体之间具有较好的公平性,此时,模型在准确率与公平性之间达到了比较理想的平衡状态,兼顾了准确性和公平性。

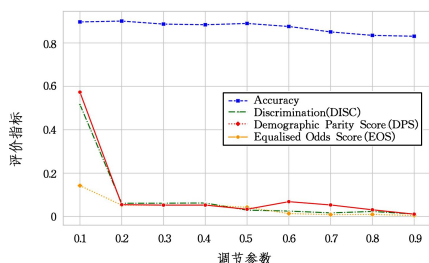


图 4 Bank 数据集上不同调节参数下模型的表现

Fig. 4 Model performance under different tuning parameters on Bank dataset

总体来看,模型在 Bank 数据集上的表现验证了其在引入公平性约束时的有效性,能够在保证分类性能的同时显著提升不同群体间的公平性,这一点与在 Statlog 和 Adult 数据集上的结果有一定的一致性。

#### 4.4.2 与普通决策树进行对比分析

为了验证本文方法在提升模型公平性方面的有效性,在 Statlog 数据集上选择最优调节参数  $\lambda = 0.8$ ,在 Adult 数据集上选择最优调节参数  $\lambda = 0.6$ ,在 Bank 数据集上选择最优调节参数  $\lambda = 0.5$ ,将这些参数应用于本文方法,作为 3 个数据集的实验参数设置。然后将此参数下训练的模型与普通决策树模型的实验效果进行对比分析。本文后续所有实验均采用这个参数设置。

由表 4 可知,本文方法在提升模型公平性方面具有明显优势,能够有效降低原始决策树模型在公平方面的偏差性,同时能够以较小的准确率损失为代价换取更优的公平性表现。

表 4 与普通决策树进行对比的结果

Table 4 Results of comparison with ordinary decision trees

指标	数据集	普通决策树	本文方法
DISC	Statlog	-0.109	<b>0.002</b>
	Adult	-0.137	<b>-0.025</b>
	Bank	0.104	<b>0.029</b>
DPS	Statlog	0.096	<b>0.050</b>
	Adult	0.119	<b>0.021</b>
	Bank	0.078	<b>0.042</b>
EOS	Statlog	0.093	<b>0.001</b>
	Adult	0.123	<b>0.041</b>
	Bank	0.111	<b>0.033</b>
准确率	Statlog	<b>0.719</b>	0.690
	Adult	<b>0.843</b>	0.770
	Bank	<b>0.901</b>	0.890

#### 4.4.3 与使用 FAIR-MDAV 处理过的数据集进行训练的其他分类器进行对比分析

在本节中,在采用 FAIR-MDAV 对原始数据进行公平性预处理后的基础上,分别使用了多种常见的分类模型(如 KNN、逻辑回归(LR)等)进行训练,并将其结果与本文方法进行比较。

根据表 5 的结果可知,在 Statlog 数据集上,本文方法在公平性指标方面的表现优于其他分类器,虽然分类准确率略低于普通决策树和逻辑回归分类器,但整体表现比其他分类器更为出色。

表 5 与其他分类器的对比结果

Table 5 Results of comparison with other classifiers

指标	数据集	KNN	LR	决策树	本文方法
DISC	Statlog	-0.033	0.073	0.030	<b>0.002</b>
	Adult	0.028	-0.049	-0.138	<b>-0.025</b>
	Bank	0.067	0.270	0.387	<b>0.029</b>
DPS	Statlog	0.086	0.069	0.075	<b>0.050</b>
	Adult	0.103	0.073	0.117	<b>0.021</b>
	Bank	0.043	0.160	0.186	<b>0.042</b>
EOS	Statlog	0.082	0.140	0.157	<b>0.001</b>
	Adult	0.074	0.045	0.120	<b>0.041</b>
	Bank	0.144	0.330	0.352	<b>0.033</b>
准确率	Statlog	0.654	0.707	<b>0.714</b>	0.690
	Adult	0.703	0.767	<b>0.846</b>	0.770
	Bank	0.876	<b>0.895</b>	0.892	0.890

在 Adult 数据集上,本文方法在公平性指标方面相较于其他分类器表现出了明显的优势。具体地,DISC 比其他模型

更接近 0, 在 DPS 和 EOS 这两个评价指标方面, 本文方法的得分为 0.021 和 0.041, 均为最低, 表明在这些模型中, 本文方法在不同群体间的预测偏差最小。同时, 尽管本文方法的准确率低于普通决策树, 但与 KNN 和逻辑回归相比, 仍然具有不错的预测性能。

在 Bank 数据集上, 本文方法同样在多个公平性指标上显著低于其他方法。具体地, 本文方法准确率低于逻辑回归和决策树分类器, 但与其他分类器相差不大, 显示出本文方法并未牺牲过多预测性能。

#### 4.4.4 与 Discrimination Aware Decision Tree Learning 方法进行对比分析

为了验证本文方法相较于 Discrimination Aware Decision Tree Learning 方法的优势, 在 Statlog, Adult, Bank 数据集上分别进行对比实验。

从图 5 可以看出, 在 Statlog 数据集上, 在公平性指标方面和分类准确度方面, 本文方法都比 Discrimination Aware Decision Tree Learning 方法表现优越。

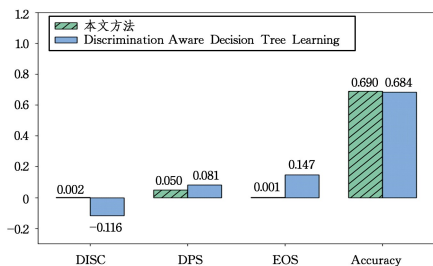


图 5 Statlog 数据集上两种方法性能比较

Fig. 5 Performance comparison of the two methods on Statlog dataset

从图 6 可以看出, 尽管 Discrimination Aware Decision Tree Learning 方法在 Adult 数据集上分类准确率较高, 但该方法在评价指标 DISC 上出现了  $-0.154$  的负值, 歧视度量越接近 0, 意味着模型对所有群体的预测越公平, 这表明该方法在公平性上存在显著偏差。相比之下, 本文提出的方法虽然分类准确率较低, 但歧视度量更接近 0, 体现了更优的公平性表现, 且在其他公平性评价指标上的表现也更优, 尽可能在公平性和准确率之间达到了较好的权衡。

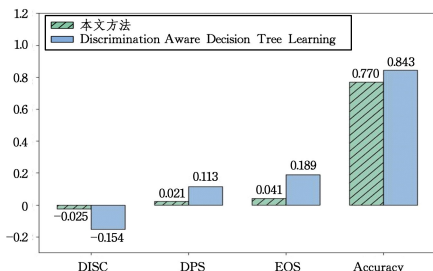


图 6 Adult 数据集上两种方法的性能比较

Fig. 6 Performance comparison of the two methods on Adult dataset

由图 7 可知, 在 Bank 数据集上, 两种方法的分类准确率比较接近, 但本文方法在公平性评价指标上的表现比 Discrimination Aware Decision Tree Learning 方法更佳, 特别是在 EOS 评价指标上, 本文方法的评价指标 EOS 值仅为 0.033, 意味着本文方法对不同群体的预测结果更公平。

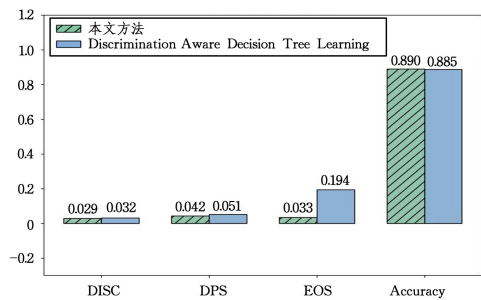


图 7 Bank 数据集上两种方法的性能比较

Fig. 7 Performance comparison of the two methods on Bank dataset

**结束语** 本文针对模型的内在偏见问题, 先将数据分成小集合, 采用更精细的方法在小集合内进行公平预处理。在构建决策树时, 本文提出了一种新型的决策树分裂标准, 综合考虑了分类准确性和公平性, 以解决传统决策树在公平性方面的缺陷, 在每次决策树分裂时, 既关注分类过程中纯度的提升, 又着力于减少不同群体间的预测偏差, 同时提出了调节参数  $\lambda$  使得模型在公平性与分类准确性之间可以灵活调整。实验结果表明, 在提升模型公平性的同时, 本文方法对模型准确率的影响较小, 且在降低歧视方面比传统方法更有效。在未来的工作中, 可以进一步探索将本文技术应用到更复杂的模型和数据集的可能性, 与其他公平性增强技术结合, 以实现更全面的公平性保障。

## 参考文献

- [1] PESSACH D, SHMUELI E. A review on fairness in machine learning[J]. ACM Computing Surveys, 2022, 55(3): 1-44.
- [2] KANAKIS M E, KHALILI R, WANG L. Machine learning for computer systems and networking: A survey[J]. ACM Computing Surveys, 2022, 55(4): 1-36.
- [3] CATON S, HAAS C. Fairness in machine learning: A survey[J]. ACM Computing Surveys, 2024, 56(7): 1-38.
- [4] CALDERS T, VERWER S. Three naive bayes approaches for discrimination-free classification[J]. Data Mining and Knowledge Discovery, 2010, 21: 277-292.
- [5] KAMIRAN F, CALDERS T. Data preprocessing techniques for classification without discrimination[J]. Knowledge and Information Systems, 2012, 33(1): 1-33.
- [6] CALDERS T, KAMIRAN F, PECHENIZKIY M. Building classifiers with independency constraints[C] // 2009 IEEE International Conference on Data Mining Workshops. IEEE, 2009: 13-18.
- [7] ZHANG W, TANG X, WANG J. On fairness-aware learning for non-discriminative decision-making [C] // 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, 2019: 1072-1079.
- [8] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [9] FRIEDLER S A, SCHEIDEGGER C, VENKATASUBRAMANIAN S, et al. A comparative study of fairness-enhancing interventions in machine learning[C] // Proceedings of the Confe-

- rence on Fairness, Accountability, and Transparency. 2019: 329-338.
- [10] TAE K H, ROH Y, OH Y H, et al. Data cleaning for accurate, fair, and robust models; A big data-AI integration approach [C]//Proceedings of the 3rd International Workshop on Data Management for End-to-end Machine Learning. 2019:1-4.
- [11] GONZÁLEZ-ZELAYA V, SALAS J, MEGÍAS D, et al. Fair and private data preprocessing through microaggregation[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 18(3): 1-24.
- [12] ZAFAR M B, VALERA I, ROGRIGUEZM G, et al. Fairness constraints; Mechanisms for fair classification[C]//Artificial Intelligence and Statistics. PMLR, 2017:962-970.
- [13] LE QUY T, ROY A, IOSIFIDIS V, et al. A survey on datasets for fairness-aware machine learning[J]. Wiley Interdisciplinary Reviews; Data Mining and Knowledge Discovery, 2022, 12(3): e1452.
- [14] HARDT M, PRICE E, SREBRO N. Equality of opportunity in supervised learning[J]. Advances in Neural Information Processing Systems, 2016, 29:3323-3331.
- [15] BERK R, HEIDARI H, JABBARI S, et al. Fairness in criminal justice risk assessments; The state of the art[J]. Sociological Methods & Research, 2021, 50(1):3-44.
- [16] ZAFAR M B, VALERA I, GOMEZ RODRIGUEZ M, et al. Fairness beyond disparate treatment & disparate impact; Learning classification without disparate mistreatment[C]//Proceedings of the 26th International Conference on World Wide Web. 2017:1171-1180.
- [17] LI Y, SUN H, WANG W H. Towards fair truth discovery from biased crowdsourced answers [C] // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020:599-607.
- [18] KAMIRAN F, CALDERS T, PECHENIZKIY M. Discrimination aware decision tree learning[C]//2010 IEEE International Conference on Data Mining. IEEE, 2010:869-874.
- [19] SPINELLI I, SCARDAPANE S, HUSSAIN A, et al. Fairdrop: Biased edge dropout for enhancing fairness in graph representation learning[J]. IEEE Transactions on Artificial Intelligence, 2021, 3(3):344-354.
- [20] TANGIRALA S. Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm[J]. International Journal of Advanced Computer Science and Applications, 2020, 11(2):612-619.



**JIANG Wenhui**, born in 2000, postgraduate. Her main research interests include machine learning and so on.



**YE Jianhong**, born in 1976, Ph.D, associate professor, master supervisor. His main research interests include process mining and so on.