

代价敏感学习方法综述*

万建武¹, 杨明²

¹(常州大学 信息科学与工程学院, 江苏 常州 213164)

²(南京师范大学 计算机科学与技术学院, 江苏 南京 210023)

通讯作者: 杨明, E-mail: myang@njnu.edu.cn



摘要: 分类是机器学习的重要任务之一.传统的分类学习算法追求最低的分类错误率,假设不同类型的错误分类具有相等的损失.然而,在诸如人脸识别门禁系统、软件缺陷预测、多标记学习等应用领域中,不同类型的错误分类所导致的损失差异较大.这要求学习算法对可能导致高错分损失的样本加以重点关注,使得学习模型的整体错分损失最小.为解决该问题,代价敏感学习方法引起了研究者的极大关注.以代价敏感学习方法的理论基础作为切入点,系统阐述了代价敏感学习的主要模型方法以及代表性的应用领域.最后,讨论并展望了未来可能的研究趋势.

关键词: 代价敏感;损失;分类;人脸识别;软件缺陷预测;多标记学习

中图分类号: TP18

中文引用格式: 万建武,杨明.代价敏感学习方法综述.软件学报,2020,31(1):113-136. <http://www.jos.org.cn/1000-9825/5871.htm>

英文引用格式: Wan JW, Yang M. Survey on cost-sensitive learning method. Ruan Jian Xue Bao/Journal of Software, 2020,31(1): 113-136 (in Chinese). <http://www.jos.org.cn/1000-9825/5871.htm>

Survey on Cost-sensitive Learning Method

WAN Jian-Wu¹, YANG Ming²

¹(School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

²(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: Classification is one of the most important tasks in machine learning. Conventional classification methods aim to attain low recognition error rate and assume the same loss from different kinds of misclassifications. However, in the applications such as the door-locker system based on face recognition, software defect prediction and multi-label learning, different kinds of misclassification will lead to different losses. This requires the learning methods to pay more attention to the samples with high-cost misclassification, and thus make the total misclassification losses minimized. To deal with this issue, cost-sensitive learning has received the considerable attention from the researchers. This study takes the theoretical foundation of cost-sensitive learning as the focal point to analyze and survey its main models and the typical applications. At last, the difficulty and probable development trend of cost-sensitive learning are discussed.

Key words: cost-sensitive; loss; classification; face recognition; software defect prediction; multi-label learning

代价敏感学习是机器学习领域的重要研究内容之一.假设代价矩阵 C 已知,代价敏感学习根据最小期望损失准则^[1,2],将测试样本 x 分类为第 j 类,如果满足:

$$\phi^*(x) = \min_j \text{loss}(x, j) \quad (1)$$

其中, $\text{loss}(x, j) = \sum_i p(i|x)C_{ij}$ 表示将样本 x 分类为第 j 类的期望损失, $p(i|x)$ 表示样本属于第 i 类的后验概率, C_{ij}

* 基金项目: 国家自然科学基金(61502058, 61876087)

Foundation item: National Natural Science Foundation of China (61502058, 61876087)

收稿时间: 2018-11-24; 修改时间: 2019-04-19; 采用时间: 2019-05-17; jos 在线出版时间: 2019-08-09

CNKI 网络优先出版: 2019-08-12 12:07:56, <http://kns.cnki.net/kcms/detail/11.2560.TP.20190812.1207.001.html>

表示将第 i 类样本分类为第 j 类的错分代价。值得注意的是:当公式(1)中代价矩阵 C 的所有元素值均为 1 时,也即 $C_{ij}=1$ 对任意的 i 和 j ,代价敏感学习将退化为追求最低分类错误率的传统分类学习算法。

代价敏感学习问题广泛存在于实际应用中。例如,基于人脸识别的门禁系统^[3]希望能将入侵者拒之门外,而允许合法者进入。这里,门禁系统往往更关注对入侵者类的识别率,尤其是安全等级要求较高的场所。这是因为将入侵者识别为合法者将会引起严重的安全问题;而将合法者错分成入侵者的损失则相对较小,即便可能会对误判的合法者带来困扰。再如,软件缺陷预测^[4]希望根据软件历史开发数据,借助机器学习方法来准确预测项目中的缺陷数、类型以及出错位置等信息。在缺陷预测的过程中,一般存在两种类型的错误分类:(1) 错误拒绝,即将正确的软件模块预测为有缺陷的软件模块;(2) 错误接收,即将有缺陷的软件模块预测为正确的软件模块。很显然,一个错误接收类型的错误分类可能导致整个软件项目失败,其损失远大于错误拒绝类型的错误分类。

本文围绕代价敏感学习方法,主要综述了代价敏感学习的理论基础、应用领域、主要模型与方法以及未来可能的研究趋势。具体地,第 1 节介绍代价敏感学习的理论基础,第 2 节介绍代价敏感学习相关的代表性应用领域,第 3 节系统介绍代价敏感学习的主要模型与方法,第 4 节分析并讨论未来可能的研究趋势与待解决的难点问题。最后,第 5 节总结全文。

1 理论基础

1.1 代价的定义

在代价敏感学习中,代价的定义问题是首要解决的问题之一。事实上,早在 2000 年,Turney 就讨论了数据挖掘和机器学习领域中可能存在的各种不同类型的代价^[5],例如错分代价^[3]、数据信息获取代价^[6-11]、主动学习代价^[2,5]、计算代价^[2,5]、存储代价^[2,5]、人-机交互代价^[12]等。从 Turney 对代价的定义不难发现,代价贯穿了模型学习的整个过程。这是因为根据“没有免费午餐定理”^[13],学习模型在获得某些方面性能提升的同时,必然以牺牲其他方面的性能为代价。例如,监督学习算法在获取强判别能力的同时必然需要付出较高的信息获取代价,用于标注样本的监督信息。因此,可定义不同类型的代价,用于度量学习模型在建模过程中的各类损失,从而获得满足最小期望损失准则的学习模型,如公式(1)所示。这里,主要关注目前研究最为广泛的一种代价——错分代价。

为了不失一般性,假设两类分类问题,其中,正负类样本分别用符号“1”和“0”表示。代价敏感学习通过定义错分代价,度量由机器学习算法错误分类所导致的不等错分损失。具体地,错分代价矩阵可定义为如下形式(见表 1)。

Table 1 Misclassification cost matrix C for the two-class classification problem

表 1 两类分类问题的错分代价矩阵 C

	真实类别为负类	真实类别为正类
预测为负类	C_{00}	C_{10}
预测为正类	C_{01}	C_{11}

其中, C_{ij} 表示将第 i 类样本分类为第 j 类的错分代价,其值越大,表明该错误分类所导致的损失越大。在文献[1]中,研究者提出了所谓的“合理性条件”,即,认为由错误分类所导致的损失应该大于正确分类的损失。因此,代价间存在如下的大小关系: $C_{01} > C_{00}, C_{10} > C_{11}$ 。

在代价敏感学习中,代价矩阵 C 可由领域专家根据经验事先设定或者采用参数学习的方法获得其值。而且,一般而言,对于不同的代价敏感应用问题,其代价矩阵不同,需结合具体的应用问题分别加以设定。根据文献[1,14],代价矩阵 C 一般应满足如下性质。

- 代价矩阵中某一行的值不能全部大于另一行的值。这是因为对于代价矩阵的第 m 和 n 行,如果在所有的列上都有 $C_{mj} > C_{nj}$,那么根据公式(1)所定义的最小期望损失准则,测试样本的预测类别恒为第 n 类。一般而言,这样的代价矩阵是不合理的。

进一步,为了减少代价矩阵 C 中的变量数量,降低代价矩阵的设置难度,研究者认为:在不改变代价敏感学习

算法最优解的前提下,可对代价矩阵进行加、减、乘、除运算.例如,根据 Elkan 提出的“合理性条件”可知,表 1 中 C_{00} 和 C_{11} 的值最小,故可将表 1 中的第 1 列和第 2 列元素分别减去 C_{00} 和 C_{11} ,得到如表 2 所示的简化代价矩阵 C' .

Table 2 Simplified misclassification cost matrix C' for the two-class classification problem

表 2 两类分类问题的简化错分代价矩阵 C'

	真实类别为负类	真实类别为正类
预测为负类	0	C'_{10}
预测为正类	C'_{01}	0

其中, $C'_{01} = C_{01} - C_{00}$, $C'_{10} = C_{10} - C_{11}$. 这里, $C'_{00} = C'_{11} = 0$ 还可解释为正确分类不会带来损失,因此其错分代价为 0. 基于此,对于两类分类问题,研究者常定义如表 2 所示的代价矩阵. 在本文后续的内容中,若不作特别说明,两类分类问题的错分代价矩阵 C 都采用表 2 的定义.

值得注意的是:表 1 所示的错分代价矩阵属于类依赖代价,即,假设将同一类中任意样本错分成另外一类的代价均相同. 此外,还存在样本依赖代价^[15-17],即,将同一类中样本错分成另外一类的代价均不相同. 例如:在银行放贷问题中,对于一笔金额为 x 美元的借贷申请,如果将钱放贷给诈骗者,银行将血本无归,损失全部 x 美元. 所以,同属于诈骗者类,但由于借贷金额 x 的不同,会产生不同的错分代价.

1.2 代价敏感的分类决策

对于两类分类问题,假设不同错误分类的损失相同,即令 $C_{00}=C_{10}=C_{01}=C_{11}=1$,由贝叶斯公式^[18]可知:如果样本 x 属于正类的概率 $p(1|x) > 0.5$,则将该样本分类为正类. 故在 C4.5 分类器等许多传统分类模型中,往往以 0.5 作为分类器的决策阈值^[18].

这里考虑代价敏感学习问题,由公式(1)可知:如果要将样本 x 分类为正类,当且仅当:

$$p(1|x)C_{11} + p(0|x)C_{01} < p(1|x)C_{10} + p(0|x)C_{00} \tag{2}$$

由表 2 所定义的代价矩阵可知 $C_{11}=C_{00}=0$,故公式(2)可简写为

$$p(0|x)C_{01} < p(1|x)C_{10} \tag{3}$$

又因为 $p(0|x) + p(1|x) = 1$,从而可得:

$$p(1|x) > p^* = \frac{C_{01}}{C_{10} + C_{01}} \tag{4}$$

在公式(4)中, $p^* = C_{01}/(C_{10} + C_{01})$ 为代价敏感的分类决策阈值. 由于 C_{10} 和 C_{01} 不相等,导致 $p^* \neq 0.5$. 这使得传统的分类模型对代价敏感学习问题失效,无法准确预测未知的测试样本.

2 应用领域

本节主要介绍代价敏感学习相关的代表性应用领域,具体包括类别不平衡问题、人脸识别问题、多标记学习问题以及软件缺陷预测问题等.

2.1 类别不平衡问题

类别不平衡问题是指分类任务中不同类别的训练样本数目差别很大的情况^[13]. 假设在两类分类问题中存在 n_0 个负类样本和 n_1 个正类样本,且 $n_1 \ll n_0$,即正类样本为少数类. 根据贝叶斯公式,可得样本 x 属于第 k 类的后验概率 $p(k|x)$ 为

$$p(k|x) = \frac{p(k)p(x|k)}{p(x)} \tag{5}$$

其中, $k=0,1$, $p(k)$ 为第 k 类的类先验概率, $p(x|k)$ 为第 k 类的概率密度函数. 这里假设正负类样本的概率密度函数相同,即 $p(x|0) = p(x|1)$. 进一步,可得:

$$\frac{p(1|x)}{p(0|x)} = \frac{n_1}{n_0} \quad (6)$$

又因为 $p(0|x)+p(1|x)=1$,公式(6)可改写为

$$p(1|x) = \frac{n_1}{n_0 + n_1} \quad (7)$$

对于类别不平衡问题,公式(7)定义了将样本分为正类也即少数类时的决策阈值 $n_1/(n_0+n_1)$.由于正类样本数远少于负类样本数 $n_1 \ll n_0$,故决策阈值 $n_1/(n_0+n_1) \ll 0.5$.

如第 1 节所述,传统的分类学习算法在两类分类问题上,通常以 0.5 作为决策阈值.很显然,在类别不平衡问题中,以 0.5 作为决策阈值并不合适,可能导致分类器对多数类具有分类倾向性,无法准确预测出少数类样本.这里,为了准确识别少数类样本,可采用代价敏感学习的方法,将少数类视为重要类别,并令其错分代价 C_{10} 大于多数类的错分代价 C_{01} .

2.2 人脸识别问题

假设在人脸识别系统中,前 $c-1$ 类为 gallery,用符号 $G_i, i=1,2,\dots,c-1$ 表示;第 c 类为 imposter,用符号 I 表示.对于任意一个测试样本,人脸识别系统可能存在 3 种类型的错分^[3]:(1) 错误拒绝,即将 gallery 错误识别为 imposter;(2) 错误接收,即将 imposter 错误识别为 gallery;(3) 错误鉴别,即将某个 gallery 错误识别为其他 gallery.

为了度量这 3 种错误分类所引起的不等错分损失问题,可采用代价敏感学习的方法,定义其错分代价.具体地,令 C_{ij} 表示将第 i 类样本分类为第 j 类的错分代价.对于不同的 i 和 j ,可得:

$$C_{ij} = \begin{cases} 0, & \text{if } i = j \\ C_{GG}, & \text{if } i < c, j < c \text{ and } i \neq j \\ C_{GI}, & \text{if } i < c \text{ and } j = c \\ C_{IG}, & \text{if } i = c \text{ and } j < c \end{cases} \quad (8)$$

其中, C_{GG} 、 C_{GI} 、 C_{IG} 分别表示错误鉴别、错误拒绝和错误接收的代价.这里假设正确识别的代价为 0,故对任意 $i=j, C_{ij}=0$.进一步,根据公式(8)可得面向人脸识别问题的错分代价矩阵 C (见表 3).

Table 3 The misclassification cost matrix for face recognition problem

表 3 面向人脸识别问题的错分代价矩阵

	第 1 个 gallery	...	第 $c-1$ 个 gallery	Imposter
预测为第 1 个 gallery	0	...	C_{GG}	C_{IG}
...
预测为第 $c-1$ 个 gallery	C_{GG}	...	0	C_{IG}
预测为 imposter	C_{GI}	...	C_{GI}	0

值得注意的是:可通过对代价 C_{GG} 、 C_{GI} 与 C_{IG} 设置不同的值,以解决不同的人脸识别应用问题.具体地:

- 在基于人脸识别的门禁系统中^[3],将 gallery 和 imposter 分别视为合法者类和入侵者类.门禁系统的任务在于准确地将入侵者拒之门外,而允许合法者进入.明显地,错误接收带来的损失要远高于错误拒绝,而错误鉴别的损失最低.因此,可通过设置 C_{GG} 、 C_{GI} 与 C_{IG} 的值,使其满足 $C_{IG} > C_{GI} > C_{GG}$;
- 在基于人脸识别的罪犯排查系统中^[3],其任务在于从人脸库中准确辨别罪犯,从而剔除合法公民的图像,保留疑似罪犯的图像,以便公安机关进一步判别.这里,错误拒绝类型的错分将带来严重的社会安全问题,因为罪犯逃过了检验,其图像已从人脸库中剔除.针对该问题,可通过设置 C_{GG} 、 C_{GI} 与 C_{IG} 的值,使其满足 $C_{GI} > C_{IG} > C_{GG}$.

此外,在人脸识别问题中,还存在类别不平衡问题和不同的类数据分布密度问题^[19].例如,在基于人脸识别的门禁系统中存在着两种不同的类别不平衡问题:(1) imposter 类的样本数远小于所有 $c-1$ 个 gallery 类的样本数;(2) imposter 类的样本数和某一个 gallery 类的样本数也一般不相同.进一步,由于 imposter 类是由不同的 imposter 构成的混合类,其类数据分布和单个 gallery 类的数据分布显著不同.

综上,人脸识别是一个复杂的代价敏感学习问题,同时存在不同错分损失问题、类别不平衡问题以及不同类数据分布问题.

2.3 多标记学习问题

多标记学习是机器学习和模式识别领域的热点研究问题^[20].多标记学习假设样本可能存在多个标记,且标记之间存在相关性.目前,多标记学习已从最初的多主题文本分类领域,逐渐发展到自动图像标注、自动音频标注、蛋白质功能预测、场景分类等领域.

近年来,有学者将多标记学习视为一种代价敏感学习问题^[21-24].这是因为在多标记数据集中广泛存在着:(1) 单个样本上的噪声标记问题;(2) 单个标记上的类别不平衡问题;以及(3) 标记间内在的代价敏感问题.具体如图 1 所示,其中, y_{ij} 表示样本 j 的第 i 个标记, n 为总的样本数, c 为可能的标记数.

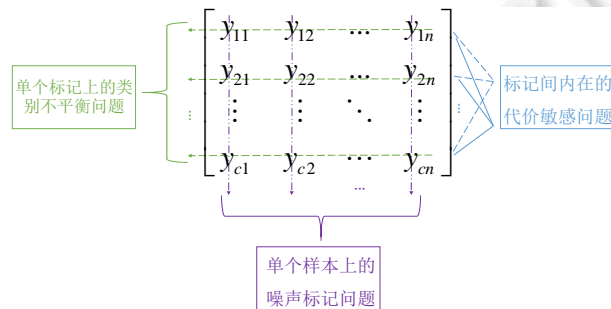


Fig.1 The cost-sensitive problems in multi-label learning

图 1 多标记学习中的代价敏感学习问题

- 单个样本上的噪声标记问题:是指样本所隶属的多个类别标记中可能存在错误标记问题.例如,在自动音乐标注等领域^[21,22],由于受主观因素的影响,包括音乐标注人音乐背景的差异、音乐标注人音乐偏好的不同,导致对同一段音乐的理解存在偏差,以致标注的信息可能不准确,故而存在噪声标记问题,严重时甚至某个标记上的所有样本标记都为噪声标记;
- 单个标记上的类别不平衡:是指在单个类别标记上,具有该标记的样本数远小于不具有该标记的样本数.例如,开源社区软件标注等问题中^[24],由于样本的候选标记集合较大,导致在单个标记上具有该标记的样本数远少于不具有该标记的样本数,这就在单个标记上形成了所谓的类别不平衡问题;
- 标记间内在的代价敏感问题:是指在样本所隶属的多个类别标记中,某些类别标记的重要性自然高于其他类别标记.这里,类别标记的重要性不是因为标记间存在的标记相关性或其他因素所导致,而是由多标记应用问题本身所导致的^[23].例如,在计算机辅助疾病诊断中,一个患者可能同时具有“感冒”“发热”“肺炎”等多个标记信息.一般宁可误检,而不允许漏检.另外,在标记预测过程中将“肺炎”标记误检的损失较大,“发热”标记次之,误检“感冒”标记的损失最小.这是因为:如果将“肺炎”标记误检,患者没有及时治疗,可能会导致呼吸困难、高烧,甚至导致患者的死亡.

2.4 软件缺陷预测问题

为了提高软件质量,减少软件中存在的缺陷,避免软件故障发生,许多研究者采用机器学习方法,用以快速、高效地判断软件各模块是否含有缺陷以及标定缺陷的具体位置^[4,25].对于软件缺陷预测问题,研究者最初关注如何提取有效的软件特征,用以训练软件缺陷预测模型.代表性的软件特征包括代码行数特征、基于控制流图的 McCabe 特征^[26]、基于模块操作数和操作符的 Halstead 特征^[27]以及考虑多态性和封装性的 CK 特征^[28]等.

近些年的研究表明:在软件项目中正确的软件模块占 80%,只有 20%的软件模块含有缺陷^[29].因此,软件缺陷预测中存在类别不平衡问题.此外,将有缺陷软件模块预测为无缺陷模块,可能导致整个软件项目的失败.尤其是在卫星发射等重要领域,软件项目的失败将会导致数十亿美元的经济损失.而将无缺陷软件模块错误

识别为缺陷模块,虽然会导致误报,影响其预测精度,但对整个软件项目而言则几乎没有损失.可见,软件缺陷预测不仅是一个类别不平衡问题,同时也是一个代价敏感学习问题^[30-33].对于该问题,可通过定义如表 4 所示的错分代价矩阵,度量相应的错分损失.这里,分别用符号“1”和“0”表示有缺陷和无缺陷的软件模块.

Table 4 The misclassification cost matrix C for software defect prediction problem

表 4 面向软件缺陷预测问题的错分代价矩阵 C

	缺陷模块	无缺陷模块
预测为缺陷模块	0	C_{01}
预测为无缺陷模块	C_{10}	0

3 主要的模型与方法

本节介绍代价敏感学习相关的主要模型与方法,用以解决类别不平衡问题、人脸识别问题、多标记学习问题以及软件缺陷预测等实际应用问题.在此之前,首先回顾分类学习模型的训练与测试过程.如图 2 所示,主要包括训练数据输入、模型训练以及结果预测这 3 个部分.这里,可根据分类的上述 3 个过程,将现有的代价敏感学习方法分为如下 3 类方法^[34]:数据前处理方法(pre-processes the training data)、结果后处理方法(post-processes the output)以及直接的代价敏感学习方法(direct cost-sensitive learning).

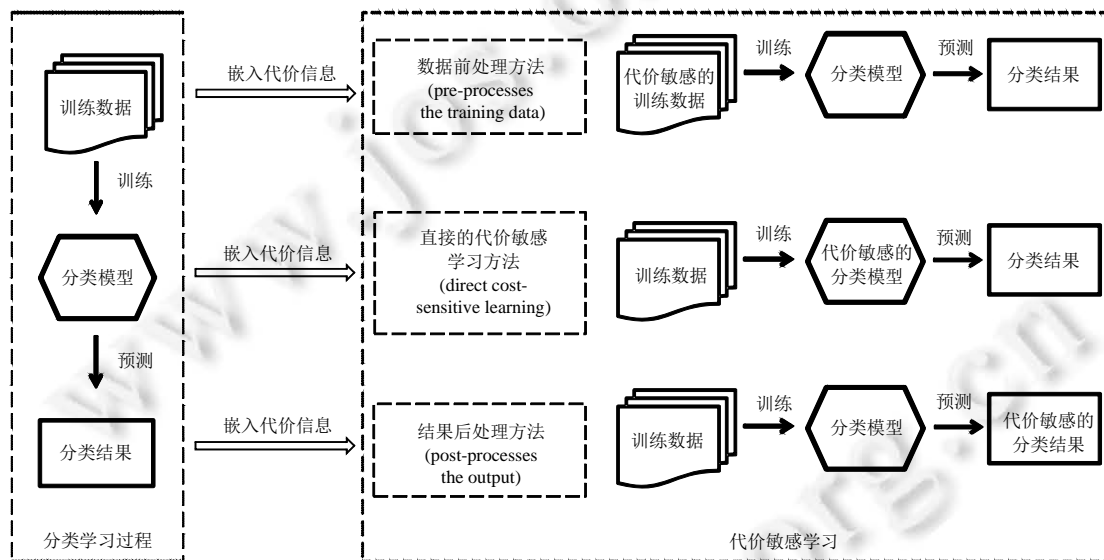


Fig.2 The classification process and the cost-sensitive learning process

图 2 分类学习过程和代价敏感学习过程

3.1 数据前处理方法

如图 2 所示,数据前处理方法旨在通过修改原始数据集,使得在新数据集上进行分类决策的结果等价于在原数据集上采用如公式(4)所示的代价敏感分类决策.根据不同的数据修改策略,可将数据前处理方法分类为:采样方法和加权方法.下面分别介绍这两类方法.

3.1.1 采样方法

采样方法通过修改原始数据集的数据分布,使得修改后的新数据集代价敏感,从而可以有效解决代价敏感学习问题.下面主要介绍两种最具代表性的采样方法.

(1) Rebalancing 方法

Rebalancing 方法^[1]通过对正负类样本进行采样,从而实现原数据集数据分布的修改.具体地,令 p' 和 p 分

别表示分类器在原数据集和修改后的数据集上的决策阈值.根据文献[1]中的定理 1 可知,原数据集中的负类样本,需按如下的系数 $ratio$ 进行采样,才能使得按 p 决策阈值对新数据集分类和以 p' 决策阈值对原数据集分类的结果一致:

$$ratio = \frac{p'}{1-p'} \frac{1-p}{p} \quad (9)$$

根据公式(9)所示的结论,Rebalancing 方法解决代价敏感学习问题的一般步骤可归纳为:

- 根据第 2 节所述的方法设定代价矩阵 C ,并按公式(4)计算代价敏感的决策阈值 $p'=p^*$;
- 任选一种代价不敏感的分类器 $f(x)$,假设其决策阈值为 p ;
- 根据公式(9)调整正负类样本的比例.具体地,假设原数据集 D 中包含有 n_0 个负类样本和 n_1 个正类样本,调整后的数据集 D' 中应有 $n_0 \times ratio$ 个负类样本和 n_1 个正类样本;
- 根据调整后的数据集 D' 训练分类器 $f(x)$,并按 p 决策阈值分类.

(2) Costing 方法

为了避免一次采样的随机性,Rebalancing 采用多次有放回的采样策略来调整数据集中的负类样本数.但这有可能导致采样后的数据集中存在重复样本,使得基于其上的训练模型出现过拟合.为解决该问题,Costing 方法^[35]提出了一种拒绝采样(rejection sampling)的策略,其基本思想是:通过对原数据集 D ,采用无放回采样策略,并根据预设的接收概率 C_{ij}/Z ,将第 i 类样本放入采样集合 D' 中.这里, Z 为常数.上述过程重复多次,并在每一个采样集合 D' 上训练子分类器 $f_i(x)$.最后,将多个子分类器集成获得最终的分类模型 $f(x) = \text{sign}(\sum_i f_i(x))$.

值得注意的是:在采样方法中,可选择对少数类采用过采样(over-sampling)^[36,37]方法或者对多数类采用欠采样方法(under-sampling)^[36,37]来调整数据集中正负类样本的比例.对这两种采样方法的性能,有学者进行了大量的实验分析和对比,实验结果表明:在代价敏感学习问题中使用欠采样方法,可取得较优的性能^[36,37].

3.1.2 加权方法

采样方法改变了原数据集的数据分布,可能会影响后续学习算法的性能.不同于采样方法,加权方法希望在不改变数据分布的前提下,通过对数据集中的样本赋予不等的权重,以达到和调整数据分布相同的分类效果.

(1) Rescaling 方法

假定代价敏感学习问题的代价矩阵 C 已知,Rescaling 方法^[1,38,39]首先给训练集中每一类样本赋予正比于其错分代价的权重,然后训练分类器进行模型预测.这里,常采用 C4.5 分类器^[18,40]是因为将样本权重嵌入熵的计算过程较为简便.具体地,令 w_1 和 w_0 分别表示正负类样本的权重,Rescaling 方法对正负类样本的加权重比

$$\frac{w_1}{w_0} = \frac{C_{10}}{C_{01}} \quad (10)$$

进一步,对于多分类问题,由 Ting 提出的代价敏感决策树算法^[41]可知,第 k 类样本的加权重 w_k 可表示为

$$w_k = h(k) \frac{n}{\sum_j h(j)n_j} \quad (11)$$

其中,重要性函数 $h(k) = \sum_j C_{kj}$ 表示第 k 类的重要性, n 和 n_j 分别表示数据集中总样本数和第 j 类的样本数.

(2) Rescale_{new} 方法

Zhou 等人研究发现^[38,39]:对于两类分类问题,无论是采用 Rescaling 方法对样本加权,还是采用 Rebalancing 方法调整原数据集的数据分布,它们对分类结果的影响是等效的.这是因为根据公式(10),容易得到 Rescaling 方法对正负类样本的加权重比 $w_1/w_0 = C_{10}/C_{01}$.

进一步,对于 Rebalancing 方法,如果假设所采用分类器的决策阈值为 $p=0.5$,则公式(9)可改写成:

$$ratio' = \frac{p^*}{1-p^*} = \frac{C_{01}}{C_{10}} \quad (12)$$

从而可以得到采样后正负类样本数的比值:

$$\frac{n'_1}{n'_0} = \frac{n_1}{n_0 \times ratio'} = \frac{C_{10} n_1}{C_{01} n_0} = \frac{w_1 n_1}{w_0 n_0} \quad (13)$$

明显地,如果选用相同的分类器用于后续模型训练和样本预测,基于公式(13)对正负类样本进行采样,或者采用 $w_1/w_0=C_{10}/C_{01}$ 对正负类样本加权,它们的分类结果都是一致的。

然而,通过进一步分析公式(11)可知:在一个含有 c 个类别的多类分类问题中,如果采用 Rescaling 方法分别对第 i 和 j 类中的样本进行加权,其加权系数可表示为

$$\frac{w_i}{w_j} = \frac{h(i)}{h(j)} = \frac{\sum_k C_{ik}}{\sum_k C_{jk}} \quad (14)$$

明显地,这时 $w_i/w_j \neq C_{ij}/C_{ji}$. 由此,Zhou 等人认为,Rescaling 和 Rebalancing 方法在多类分类问题上不等价。

为了解决该问题,Zhou 等人提出了 Rescale_{new} 方法^[38,39],它通过将多类分类问题分解为 $\binom{2}{c}$ 对两类分类问题,并且约束在每一个两类分类问题中,其加权的权重系数都需满足 $w_i/w_j=C_{ij}/C_{ji}$. 通过联立求解方程组,希望获得一组满足所有约束要求的加权系数 $[w_1, w_2, \dots, w_c]$.

(3) 同时处理代价敏感学习问题和类别不平衡问题的 Rescaling 方法

在一些实际应用问题中,数据集中可能同时存在代价敏感学习问题和类别不平衡问题.例如,在软件缺陷预测中,正确的软件模块占 80%,只有 20% 的软件模块含有缺陷,即 $n_0 \gg n_1$,其中, n_1 和 n_0 分别表示正负类的样本数.这里默认将少数类视为正类,多数类为负类.另一方面,将缺陷模块预测为无缺陷模块可能会导致整个软件项目的失败,其损失远高于将无缺陷模块预测为缺陷模块所带来的损失,即 $C_{10} \gg C_{01}$.

为了解决代价敏感学习问题,可采用公式(10)所示的加权系数 $w_1/w_0=C_{10}/C_{01}$,分别给正负类样本加权.进一步,为了解决类别不平衡问题,使得训练集中正负类样本数均衡,可根据公式(9)分别给正负类样本加权.具体地,假设所采用分类器的决策阈值为 $p=0.5, C_{10}=n_0, C_{01}=n_1$,从而可得加权系数 $ratio''=n_1/n_0$.

综上,可得同时处理代价敏感学习问题和类别不平衡问题的统一加权系数^[42]:

$$r = \frac{w_1}{w_0} \frac{1}{ratio''} = \frac{C_{10} n_0}{C_{01} n_1} \quad (15)$$

3.1.3 分析与小结

为解决代价敏感学习问题,采样方法采用修改数据分布的策略,但数据分布的改变可能会影响后续分类模型的构建,使其无法有效挖掘数据之间的潜在关系.加权方法在不改变数据分布的前提下,通过采用样本加权的策略,可有效解决两类分类问题.但对于多类分类问题,如何学得一组满足所有约束要求的加权系数 $[w_1, w_2, \dots, w_c]$ 是加权方法研究的难点问题.例如在 Rescale_{new} 方法中,其每一类样本的加权系数是通过联立求解方程组而获得的.当方程组的系数矩阵满秩时,Rescale_{new} 无法求得一组满足所有约束条件的加权系数^[38,39].此外,Rescale_{new} 方法在解决多类样本的加权问题时,通过将多类分类问题转化为多个两类分类问题,其内在假设是采用“一对一”的多类分类模式.然而,当采用“一对多”的多类分类模式时,如何设计有效的学习算法以解决多类样本的加权问题,是未来值得继续研究的一个重要研究方向。

进一步,为了分析与比较采样方法和加权方法在代价敏感学习问题上的性能,Lopez 等人采用实验的方法,分析比较了 SMOTE(synthetic minority oversampling technique)^[43]、ENN(Wilson's edited nearest neighbor rule)^[44] 等采样方法与 Ting 提出的代价敏感决策树算法^[41] 的实验性能.结果表明:在代价敏感学习问题中,无论是采样方法还是加权方法,其性能都没有显著地好于另一种方法^[45].因此,采样方法和加权方法都被认为是解决代价敏感学习问题的两种有效且值得继续研究的方法。

3.2 结果后处理方法

结果后处理方法旨在通过采用调整分类器决策阈值的策略,来解决代价敏感学习问题.结果后处理方法对所选用的分类器有一个很重要的要求,即要求分类器能够输出每一个测试样本隶属于各类的概率.例如,对于两类问题,由公式(4)可知,代价敏感的决策阈值为 $p^*=C_{01}/(C_{10}+C_{01})$. 这里,如果所采用的分类器可输出测试样本 x

隶属于正类的概率 $p(1|x)$ 且 $p(1|x) > p^*$, 那么就将该样本分类为正类. 由此可见: 对于结果后处理方法来说, 如何准确预测测试样本隶属于各类的概率是关键. 下面分别介绍几种代表性的结果后处理学习算法.

(1) MetaCost 方法

MetaCost 方法^[46]的基本思想是利用公式(1)所定义的最小期望损失准则对训练样本进行重标记, 然后在重标记的数据集上训练新的分类模型, 使其代价敏感. 具体步骤如下.

- 基于 Bagging 方法, 训练多个子分类器 $f_i(x)$;
- 根据子分类器 $f_i(x)$ 对训练集中的样本进行分类, 获得样本在该子分类器下的预测概率 $p(k|x, f_i(x))$;
- 集成分类结果, 获得概率 $p(k|x) = \frac{1}{\sum_i 1} \sum_i p(k|x, f_i(x))$;
- 根据公式(1), 对训练样本进行重新标注;
- 根据重新标注后的训练集, 训练最终的预测模型 $f(x)$.

(2) ETA 方法

ETA(empirical threshold adjusting)方法^[2,34]采用交叉验证方法, 寻找分类器的最优决策阈值. 与其他结果后处理方法的不同之处在于: ETA 只要分类器输出训练样本隶属于各类的概率排序, 而非具体的预测概率值. 具体地, 对于给定的代价敏感学习问题, 当 ETA 获得分类器对训练样本的预测概率排序之后, 采用交叉验证方法计算按各决策阈值 $T_i \in [0, 1], i=1, 2, \dots, K$ 进行分类后的整体错分损失, 并得到如图 3 所示的损失曲线. 进一步地:

- 如果损失曲线有唯一的最小值, 则选取最小整体错分损失所对应的决策阈值作为 ETA 方法的最优决策阈值. 按此规则, 在图 3(a)中, 应选取 T 作为最优决策阈值;
- 如果损失曲线中有多个最小值点, 则选取周围损失曲线变化最缓慢的极值点所对应的阈值作为 ETA 的最优决策阈值. 按此规则, 在图 3(b)中, 应选取 T_2 作为最优决策阈值.

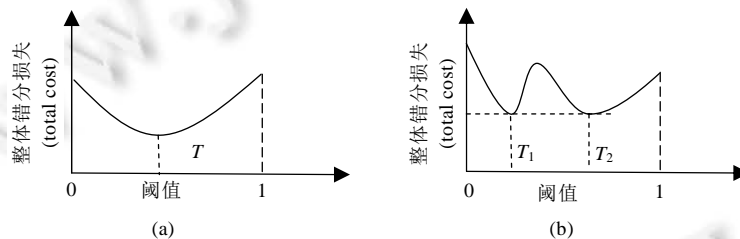


Fig.3 Overall misclassification loss vs. the decision value

图 3 整体错分损失 vs. 决策阈值

(3) Cost-sensitive Neural Networks 方法

Cost-sensitive Neural Networks^[47]采用阈值移动的方法, 即将决策阈值移向多数类, 使得少数类被错分的概率降低, 以解决类别不平衡问题. 具体地, 对于给定的测试样本 x , 采用预训练学得的神经网络模型进行分类, 可得 c 个输出函数值, 分别用 $f_k(x), k=1, 2, \dots, c$ 表示, 其值越大, 表明样本属于该类的概率越大. 这里, 为了度量每一类样本的重要性, 采用如公式(11)所示的重要性函数 $h(k) = \sum_j C_{kj}$. 明显地, 如果对于将少数类分为其他各类的错分, 设置较高的错分代价, 则少数类的 $h(k)$ 值就越大, 说明少数类比较重要. 基于此, 测试样本 x 在 Cost-sensitive Neural Networks 方法上的最终预测类别可由如下的目标函数求得:

$$\max_k : h(k) f_k(x) \quad (16)$$

3.3 直接的代价敏感学习方法

近年来, 直接的代价敏感学习方法引起了研究者的广泛关注. 与数据前处理以及数据后处理方法不同, 直接的代价敏感学习方法既不需要修改训练数据集, 也不需要调整分类器的决策阈值. 它直接将代价信息嵌入学习模型的目标函数, 通过最小化期望损失, 获得代价敏感的学习算法 f . 直接的代价敏感学习方法的一般形式为

$$\min_f \text{loss}(f, X, C) \quad (17)$$

其中, X 表示训练样本集, C 为错分代价矩阵. 这里, 可根据 f 的不同类型和不同应用领域, 将直接的代价敏感学习方法分类为: 基于不同分类模型的代价敏感分类算法、面向不同应用问题的代价敏感分类算法以及与其他学习问题相结合的代价敏感学习算法. 下面分别对这几类学习算法加以详细介绍.

3.3.1 基于不同分类模型的代价敏感分类算法

基于不同分类模型的代价敏感分类算法通过将错分代价信息直接嵌入分类模型, 获得满足最小期望损失准则的代价敏感分类算法. 具体地, 根据不同的分类模型, 可将该类方法分为如下 4 类.

(1) 代价敏感的决策树算法

由文献[48]对代价敏感决策树算法的综述可知, 目前, 研究者至少提出了 50 种以上的代价敏感决策树算法. 这些算法主要关注医学诊断等领域中获取样本信息的测试代价以及将样本错分的错分代价. 本文通过分析比较, 将现有的代价敏感决策树算法归纳为如下几类.

- 仅考虑错分代价的代价敏感决策树算法. 代表性方法包括: CSNL(cost-sensitive non-linear decision tree) 方法^[49], 它通过定义代价敏感的判别准则进行决策树的属性分裂; Bahnsen 等人提出的基于样本依赖代价的代价敏感决策算法^[17], 并将其应用于信用卡欺诈检测任务中;
- 仅考虑测试代价的代价敏感决策树算法. 代表性方法包括 Freitas 等人提出的 CS-C4.5 方法^[50]和 Davis 等人提出的 CSGain 方法^[51];
- 同时考虑测试代价和错分代价的代价敏感决策树算法. 代表性方法包括: ICET 方法^[52], 它采用遗传算法生成决策树, 这里, 在遗传算法的评价函数中综合考虑了样本的错分代价和获取样本属性所需的测试代价; 类似地, Ling 等人提出了一种测试策略用于最小化医学诊断中获取病人属性的代价和将样本错分的代价^[53]; 此外, 还有研究者基于多目标优化算法提出了一种最小化测试代价和错分代价的代价敏感决策树算法^[54]; CTS Tree(cost-time sensitive decision tree)方法^[55]通过修改决策树的属性分裂准则, 使得模型代价敏感. 这里所学习的 CTS Tree 模型倾向于选取具有高信息熵值、低等待代价和错分代价的属性; 此外, 属于该类的方法还包括文献[6-11];
- 关注过拟合问题的代价敏感决策树算法. Wang 等人基于代价敏感决策树算法提出了 3 种避免学习模型过拟合的策略^[56]: (1) 在创建决策树之前, 采用特征选择方法预处理数据; (2) 在创建决策树的过程中, 采用 smoothing 方法; (3) 基于预定义代价阈值的剪枝策略.

(2) 代价敏感的支持向量机

CS-SVM(cost-sensitive support vector machine)是一种解决两类分类问题的代价敏感学习算法^[57], 它通过在 SVM 模型的损失函数中嵌入代价信息, 使得学习模型代价敏感. 具体地, CS-SVM 模型的目标函数定义如下:

$$\min_f \frac{1}{2} \|f\|_2^2 + \gamma \sum_{i=1}^n \text{loss}(f(x_i), y_i, C) \quad (18)$$

其中, γ 为松弛变量, $y_i \in \{0, 1\}$ 表示 x_i 样本的真实标签, $y_i=1$ 表示样本属于正类, $y_i=0$ 表示样本属于负类. 嵌入代价信息的 hinge loss 函数定义如下:

$$\text{loss}(f(x_i), y_i, C) = \begin{cases} C_{10} \max\{0, 1 - f(x_i)\}, & \text{if } y_i = 1 \\ C_{01} \max\{0, 1 + f(x_i)\}, & \text{if } y_i = 0 \end{cases} \quad (19)$$

进一步, 考虑到公式(19)所定义的代价敏感损失函数与 CS-SVM 模型的最终决策规则存在不一致性, 研究者通过修改 hinge loss 函数提出了一种满足贝叶斯最优代价敏感分类界的代价敏感支持向量机模型^[58]. 此外, 为了解决多类分类问题, Lee 等人提出了面向多类分类问题的代价敏感支持向量机^[59]. 近年来, 还有研究者关注 CS-SVM 中的模型选择问题^[60].

(3) 代价敏感的回归模型

为了使得分类结果具有理论保证和可解释性, Zhou 等人基于贝叶斯风险决策理论^[1, 18, 40], 将错分代价信息嵌入核化的 Logistic 回归模型, 提出了一种面向多类分类问题的代价敏感核 Logistic 回归(multi-class cost-

sensitive kernel logistic regression,简称 McKLR)^[3],并将其应用于代价敏感人脸识别问题.具体地,为了使得代价敏感人脸识别的分类结果满足最小期望损失准则,McKLR 定义了如下的损失函数:

$$\text{loss}(f(x), x, C) = \begin{cases} \sum_{i=1, i \neq t}^{c-1} p(G_i | x) C_{GG} + p(I | x) C_{IG}, & \text{if } f(x) = G_t \\ \sum_{i=1}^{c-1} p(G_i | x) C_{GI}, & \text{if } f(x) = I \end{cases} \quad (20)$$

(4) 代价敏感的最近邻分类器

与 McKLR 类似,Zhou 等人基于贝叶斯风险决策理论^[1,18,40],将错分代价信息嵌入最近邻分类器中,提出了一种面向多类分类问题的代价敏感最近邻分类器(multi-class cost-sensitive K -nearest neighbor,简称 McKNN)^[3],并将其应用于代价敏感人脸识别问题.具体地,McKNN 首先根据最近邻分类器的特点,对后验概率 $p(G_i|x)$ 进行了重表示;然后将重表示后的后验概率 $p(G_i|x)$ 带入如公式(20)所示的损失函数中,得到如下的目标函数:

$$\text{loss}(f(x), x, C) = \begin{cases} \sum_{i=1, i \neq t}^{c-1} p(G_i) \prod_{m=1}^k p(l(x_m) | G_i) C_{GG} + p(I) \prod_{m=1}^k p(l(x_m) | I) C_{IG}, & \text{if } f(x) = G_t \\ \sum_{i=1}^{c-1} p(G_i) \prod_{m=1}^k p(l(x_m) | G_i) C_{GI}, & \text{if } f(x) = I \end{cases} \quad (21)$$

3.3.2 面向不同应用问题的代价敏感分类算法

根据文献调研,代价敏感分类算法常被应用于类别不平衡问题^[61,62]、人脸识别问题^[3,63-65]、软件缺陷预测问题^[66-71]、欺诈检测问题^[72]、垃圾邮件过滤问题^[73]以及计算机辅助疾病诊断^[74]等领域.下面介绍代价敏感分类算法在其中几种代表性应用领域的具体应用情况.

(1) 面向类别不平衡问题的代价敏感分类算法

目前,对于两类分类问题中的类别不平衡问题,研究者已经有了较成熟的解决方案.例如,采用数据前处理方法中的采样方法对数据集进行重采样,或者采用加权方法对样本进行加权.但是对于多类分类问题中的类别不平衡问题,还没有形成较完善统一的解决方案.

文献[61]研究代价敏感的 Boosting 算法,并将其应用于两类分类和多类分类情形下的类别不平衡问题.具体地,为了有效获得样本的错分代价信息,在两类分类问题中,采用经验方法加以设定;而在多类分类问题中,则采用遗传算法学习获得.文献[62]为了解决 DBN(deep belief network)^[75]方法无法准确预测少数类样本的问题,设计了一种 ECS-DBN(evolutionary cost-sensitive deep belief network)方法.它首先采用进化学习算法优化学习最优的样本错分代价,然后基于公式(1)所定义的最小期望损失准则,定义代价敏感的 DBN 模型.

(2) 面向代价敏感人脸识别问题的代价敏感分类算法

如第 2.2 节所述,人脸识别问题是代价敏感学习的代表性应用领域之一.据我们所知,mcKLR 和 mcKNN 方法^[3]是最早提出的两种用于解决代价敏感人脸识别问题的代表性算法.其主要贡献在于:(1) 给出了代价敏感人脸识别问题的完整问题描述,基于此,后续研究者开始采用代价敏感学习方法解决人脸识别问题;(2) mcKLR 和 mcKNN 的模型定义严格遵照贝叶斯风险决策理论^[1,18,40],其目标函数满足最小期望损失准则,使得分类结果具有理论保证和可解释性.

近些年来,受矩阵分解技术在图像处理领域成功应用的启发,研究者们提出了一系列基于矩阵分解技术的代价敏感人脸识别方法.下面对其中几种代表性算法分别加以介绍.

• CSDL 方法

考虑到字典学习算法在人脸识别领域中的广泛应用,Sun 等人提出了一种 CSDL(cost-sensitive dictionary learning)算法^[63],并将其应用于代价敏感的人脸识别问题.CSDL 方法的基本思想是:通过在字典学习模型中嵌入代价惩罚矩阵 Q ,以获得代价敏感的字典系数;然后采用代价不敏感的 SRC(sparse representation based classifier)分类器^[76]进行分类预测.CSDL 方法的目标函数定义如下:

$$\min_{U, V} \|X - UV\|_F^2 + \lambda_1 \|Q \odot V\|_F^2 + \lambda_2 \sum_{i=1}^n \|v_i\|_1 \quad (22)$$

其中, U 和 $V=[v_1, v_2, \dots, v_n]$ 为需要优化求解的字典和字典系数; $Q=[q_1, q_2, \dots, q_n]$ 为预定义的代价惩罚矩阵,其中,惩罚因子 $q_i^k = \delta(C_{l(x_i)l(x_k)})$ 度量了将第 $l(x_i)$ 类样本分类为第 $l(x_k)$ 类样本的错分损失. $\delta(C_{l(x_i)l(x_k)})$ 的具体定义为

$$\delta(C_{l(x_i)l(x_k)}) = \begin{cases} 1, & \text{if } C_{l(x_i)l(x_k)} = 0 \\ \sigma C_{l(x_i)l(x_k)}, & \text{if } C_{l(x_i)l(x_k)} \neq 0 \end{cases} \quad (23)$$

其中, $\sigma \geq 2$ 是常量.值得注意的是:在 CSDL 方法中,字典系数 V 的每一个元素值都要求大于 0,否则,公式(22)中的代价敏感字典学习项 $\|Q \odot V\|_F^2$ 可能没有物理意义.然而,CSDL 的目标函数往往很难满足这一条件.

- Sparse Cost-sensitive Classifier 方法

Sparse Cost-sensitive Classifier^[64]采用和 mcKLR、mcKNN 类似的思路,即严格按照公式(20)所示的 $loss(\cdot)$ 函数定义满足最小期望损失准则的目标函数.它们的不同之处在于: Sparse Cost-sensitive Classifier 根据稀疏表示分类器在每一类上的重构残差 $res_k, k=1, 2, \dots, c$ 估计样本的后验概率.这里,属于第 k 类的残差值 res_k 越小,样本属于第 k 类的后验概率越大.

- CSMF 方法

不同于 CSDL 方法在目标函数中直接约束字典系数稀疏且代价敏感,CSMF(cost-sensitive matrix factorization)^[65]采用两步学习策略,用以学得代价敏感的判别系数.CSMF 首先采用矩阵分解方法获得样本的高层语义表示 $V=[v_1, v_2, \dots, v_n]$;然后定义代价敏感的隐语义回归模型,将高层特征映射到样本的标签空间,使得样本的整体错分损失最小.上述过程采用交替迭代的方法,联合学习,以获得最优解.CSMF 的目标函数定义如下:

$$\min_{U, V, W} \|X - UV\|_F^2 + \mu \sum_{i=1}^n h(l(x_i)) \|W^T v_i - y_i\|_2^2 + \lambda R(U, V, W) \quad (24)$$

其中,第 1 项为矩阵分解项,用于学习高层的语义表示 V ;第 2 项为代价敏感的隐语义回归项,根据 CSMF 的定理 1 可知,其最优解可最小化整体错分损失;第 3 项为正项 $R(U, V, W) = \|U\|_F^2 + \|V\|_F^2 + \|W\|_F^2$,用于避免过拟合问题; W 为投影方向, μ 和 λ 为平衡参数.重要性函数 $h(l(x_i))$ 度量了第 $l(x_i)$ 类样本的重要性,其定义如下:

$$h(l(x_i)) = \begin{cases} (c-2)C_{GG} + C_{GI}, & \text{if } l(x_i) < c \\ (c-1)C_{IG}, & \text{if } l(x_i) = c \end{cases} \quad (25)$$

(3) 面向软件缺陷预测问题的代价敏感分类算法

如第 2.4 节所述,软件缺陷预测是代价敏感学习的代表性应用领域之一.近年来,研究者提出了大量基于代价敏感学习方法的软件缺陷预测分类算法,用于对同一项目内^[30,31,33,66]或跨项目间^[67-71]的软件模块进行缺陷预测.下面介绍两种代表性的软件缺陷预测分类算法.

- 用于软件缺陷预测的 Cost-sensitive Neural Network 方法

为了解决软件缺陷预测中的代价敏感学习问题和类别不平衡问题,Arar 等人提出了一种 Cost-sensitive Neural Network 方法^[33],通过定义如下的代价敏感损失函数,使得学习模型代价敏感:

$$loss = C_{01} \times FPR \times p_0 + C_{10} \times FNR \times p_1 \quad (26)$$

其中,缺陷模块为正类,用“1”表示;无缺陷模块为负类,用“0”表示. FPR 和 FNR 分别表示假正例率和假负例率, p_1 和 p_0 表示正负类样本的类先验概率.值得注意的是,该损失函数本质上是和公式(15)一致的.

- CSBNN 方法

Zheng 基于 Adaboost 模型提出了 3 种代价敏感的 Boosting 算法^[66].其中,第 1 种算法 CSBNN-TM(cost-sensitive boosting neural networks with threshold-moving)属于决策阈值调整方法,与公式(16)中的方法类似;另外两种算法 CSBNN-WU1 和 CSBNN-WU2(cost-sensitive boosting neural networks with weight-updating)则关注将代价信息嵌入 Boosting 算法的权重更新规则中,使得被错误分类的缺陷模块在下一轮更新中被赋予更高的权重.这里,代价敏感的权重函数定义如下:

$$W_{t+1}(x) = C_{y_f(x)} W_t(x) \exp(-f_t(x)y) / Z_t \quad (27)$$

其中, $y \in \{-1, 1\}$ 是样本 x 的真实标签, $y=1$ 为有缺陷模块,反之为无缺陷模块. $C_{y_f(x)} = 1$, 如果 $y=f(x)$, 其中, $f_t(x)$ 为 Adaboost 中的弱分类器. $W_{t+1}(x)$ 表示样本 x 经 $t+1$ 次更新后的权重.

3.3.3 与其他学习问题相结合的代价敏感学习算法

近年来,将代价敏感学习问题与其他机器学习问题相结合协同研究,逐渐成为一种新的研究思路.例如,将

代价敏感学习分别与多标记学习问题、降维、特征选择、半监督、主动学习、深度学习以及多核学习相结合,研究者提出了代价敏感的多标记学习算法^[21-24,77-80]、代价敏感的降维算法^[81-84]、代价敏感的特征选择算法^[30,31,85]、代价敏感半监督学习算法^[86-92]、代价敏感的主动学习算法^[93-95]、代价敏感的深度学习算法^[93,96]以及代价敏感的多核学习算法^[97].下面介绍代价敏感学习算法与其中几种代表性学习问题的具体结合情况.

(1) 与多标记学习问题相结合的代价敏感学习算法

近年来,代价敏感多标记学习问题引起了研究者的广泛关注.其中,最早研究的一个应用领域是音频标注问题.例如,MIREX 2009 音乐标注竞赛要求使用者根据听到的音乐片段(大约 10s)对其进行标注.目前,该数据集共有 2 472 个音乐片段,每个音乐片段有 45 种可能的音乐标注.有研究者认为:要准确识别给定音频可能的标注信息,不仅要考虑每一段音频上的不同标注信息,而且还要考虑每一个标注所获得的总标注次数.一般而言,获得标注次数较多的标注反映了标注者的普遍喜好;而获得标注次数较少的标注则反映了标注者个体的个性,但有时也可能是噪声标注.为了有效预测标注者对音乐的普遍喜好,研究者常将每个标注所获得的标注次数作为代价,设计代价敏感的多标签学习算法.下面具体介绍两种用于解决音频标注问题的代价敏感多标记学习算法.

• Cost-sensitive Stacking 方法

Cost-sensitive Stacking 方法^[21]的基本思想是,将一个含有 K 个标记的代价敏感多标记分类问题转化为 K 个代价敏感两分类问题.具体地,首先在每一个两分类问题上采用 CS-SVM 模型^[57],训练代价敏感的分类算法 $f_k(x), k=1,2,\dots,K$;然后,根据样本的真实标签和代价敏感的分类算法 $f_k(x)$ 学习代价不敏感的 stacking 分类器,用于最终的标注预测.

• Cost-sensitive RAKEL 方法

Cost-sensitive RAKEL^[21,22]首先基于经典的 RAKEL(random k -labelsets)方法^[98]将含有 K 个标记的多标记集划分为若干个大小为 k 的标记子集 R_m ;然后在每一个标记子集 R_m 上设计 CSLP(cost-sensitive label powerest)方法,进行有效的音频标注预测.这里,LP(label powerest)^[99]方法是通过将标记子集 R_m 中所有可能的标注组合转化为单标记下的不同类别属性,进而采用单标记学习方法解决多标记学习问题.因此,CSLP 需要将每个标注上的代价转化为标记组合的代价.这里,通过定义总代价的方法对每一种标记组合赋予代价,具体定义如下:

$$h^m(i) = \begin{cases} \sum_{j \in R_m, y_{ij}=1} C_{ij}, & \text{if } j \in R_m \text{ and } y_{ij} = 1 \\ 1, & \text{otherwise} \end{cases} \quad (28)$$

其中, $h^m(i)$ 表示样本 x_i 在 R_m 标记子集上的总代价; C_{ij} 表示 x_i 在第 j 个标记上的代价,由用户对该标记的标注次数加以度量; $y_{ij}=1$ 表示 x_i 具有第 j 个标记.这里假设样本不具有某一个标记的代价为 1,即 $y_{ij}=0$ 时, $C_{ij}=1$.

值得注意的是:除了上述两种代表性的代价敏感多标记分类算法之外,近年来研究者还提出了许多其他的代价敏感多标记学习算法和代价敏感多示例学习算法.具体可参见文献[21-24,77-80].

(2) 与降维学习问题相结合的代价敏感学习算法

代价敏感的分类算法仅将代价信息嵌入了分类器的设计部分,然而在模型训练过程中,特征学习是分类器设计的前序步骤.如果在特征学习环节存在大量代价信息的丢失,使得学得特征代价不敏感,这将极大地影响后续分类器的性能.为此,研究者提出了代价敏感的降维算法,其主要思想是:通过学习最优的投影方向 W ,使得期望损失最小化.这里用投影方向 W 替换公式(17)中的 f ,可得代价敏感降维模型的一般形式:

$$\min_W \text{loss}(W, X, C) \quad (29)$$

• CSPCA、CSLDA 和 CSLPP 方法

CSPCA(cost-sensitive principal component analysis)、CSLDA(cost-sensitive linear discriminant analysis)和 CSLPP(cost-sensitive locality preserving projections)^[81,82]是最早应用于代价敏感人脸识别问题的代价敏感降维算法,它们的基本思想是:采用投影后成对样本间的距离来度量将样本所属类错分为另一类的概率.进一步,将该错分概率乘以相应的错分代价,可得满足错分损失最小化准则的降维学习模型.具体地:

(1) CSPCA 算法的目标函数定义如下:

$$\max_W : \frac{1}{2} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n C_{l(x_i)l(x_j)} \|W^T x_i - W^T x_j\|_2^2 \quad (30)$$

(2) CSLDA 算法的目标函数定义如下:

$$\max_W : \frac{\text{tr}(W^T S_B W)}{\text{tr}(W^T S_W W)} \quad (31)$$

其中,代价敏感类间离散度矩阵 S_B 和代价敏感类内离散度矩阵 S_W 定义为

$$\begin{cases} S_B = \sum_{i=1}^c \sum_{j=1}^c C_{ij} (m_i - m_j)(m_i - m_j)^T \\ S_W = \sum_{i=1}^c h(i) \sum_{j=1}^{n_i} (x_j - m_i)(x_j - m_i)^T \end{cases} \quad (32)$$

其中,变量 n_i 和 m_i 分别表示第 i 类的样本数和类中心.重要性函数 $h(i)$ 的定义和公式(25)一致.

3) CSLPP 算法的目标函数定义如下:

$$\max_W : \sum_{i=1}^n \sum_{j=1}^n C_{l(x_i)l(x_j)} \|W^T x_i - W^T x_j\|_2^2 S_{ij} \quad (33)$$

其中, S_{ij} 为无监督 K 近邻图的权重矩阵.根据文献[100],其定义如下:

$$S_{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / \tau), & \text{if } x_i \text{ and } x_j \text{ are } K \text{ Nearest Neighbors} \\ 0, & \text{otherwise} \end{cases} \quad (34)$$

- WCSLPP 方法

在 CSLPP 模型中,其 K 近邻图采用无监督方式创建.因此,在样本 x_i 的 K 个近邻中可能包含同类样本 x_p 和异类样本 x_q .根据错分代价矩阵 C 的定义,将样本正确分类的代价为 0,而错误分类的代价大于 0,即 $C_{l(x_i)l(x_p)} = 0$, $C_{l(x_i)l(x_q)} > 0$, 故而可将公式(33)所定义的 CSLPP 模型改写为

$$\min_W : \sum_{i=1}^n \sum_{j=1, l(x_j) \neq l(x_i)}^n C_{l(x_i)l(x_j)} \|W^T x_i - W^T x_j\|_2^2 S_{ij} \quad (35)$$

由公式(35)可知:CSLPP 最小化了投影后异类样本间的距离,使得原空间近邻的异类样本近邻保持,降低了投影方向 W 的判别能力.

为了解决该问题,Wan 等人通过分析 LPP(locality preserving projections)算法^[100]得到了两个观察结论:(1) LPP 算法受类别不平衡影响;(2) LPP 算法本质上追求整体错分量的最小化.并且,进行了理论论证.随后,将错分代价嵌入 LPP 模型,提出了满足最小错分损失准则的 WCSLPP(weighted cost-sensitive locality preserving projections)算法^[83].进一步,考虑到在代价敏感的人脸识别问题中,每一个 gallery 类中的样本数远少于 imposter 类中的样本数,采用加权策略,平衡了各类样本对投影方向的贡献.WCSLPP 模型的目标函数定义如下:

$$\min_W : \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} \|W^T x_i - W^T x_j\|_2^2 S_{ij}^w \quad (36)$$

其中, S_{ij}^w 为监督 K 近邻图的权重矩阵, α_{ij} 为同时考虑代价敏感和类别不平衡问题的加权系数.

$$\begin{cases} S_{ij}^w = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / \tau), & \text{if } x_i \text{ and } x_j \text{ are } K \text{ Nearest Neighbors and } l(x_i) = l(x_j) \\ 0, & \text{otherwise} \end{cases} \\ \alpha_{ij} = \begin{cases} h(k)\gamma_k, & \text{if } l(x_i) = l(x_j) = k \\ 0, & \text{if } l(x_i) \neq l(x_j) \end{cases} \\ \gamma_k = \frac{1}{\sum_{q=1}^{n_k} \sum_{q'=1}^{n_k} \|x_q - x_{q'}\|_2^2 S_{qq'}^w} \end{cases} \quad (37)$$

其中,重要性函数 $h(k)$ 的定义和公式(25)一致, γ_k 和 n_k 分别表示第 k 类的加权系数和样本数.

- PCSCDA 方法

根据第 2.2 节对人脸识别问题的介绍可知:基于人脸识别的门禁系统中存在不同的类数据分布密度问题,即 imposter 类是由多个不同的 imposter 构成的混合类.具体如图 4 所示,imposter 类由 K 个 imposter 构成,且每一个 imposter 都有其各自的数据分布.这使得许多基于单高斯假设的判别分析算法的模型假设不成立,例如

CSLDA.为此,Wan 等人将人脸识别门禁系统视为一个子类学习问题^[101],提出了一种代价敏感的子类判别分析算法 PCSCDA(pairwise costs in SubClass discriminant analysis)^[84].

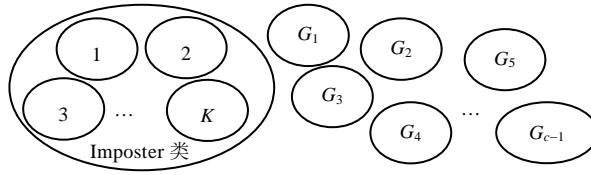


Fig.4 The data distribution of the door-locker system based on face recognition

图 4 人脸识别门禁系统的数据分布示意图

• CTKCCA 方法

在跨项目软件缺陷预测问题中,研究者希望使用来源于其他项目的软件数据训练学习模型,进而根据学得的学习模型对目标项目中的软件数据进行准确预测.类别不平衡问题以及跨项目软件数据间的异构性是跨项目软件缺陷预测研究的难点问题.为解决上述问题,CTKCCA(cost-sensitive transfer kernel canonical correlation analysis)^[68]将训练用的软件数据以及目标项目中的软件数据分别视为迁移学习中的源域和目标域,通过定义 TKCCA(transfer kernel canonical correlation analysis)模型,有效融合跨项目软件数据,解决数据异构问题;进一步,通过采用代价敏感学习方法,在 TKCCA 中有效嵌入代价信息,解决类别不平衡问题.在多种跨项目软件缺陷预测实验设置下的结果,表明了 CTKCCA 方法的有效性.

(3) 与特征选择问题相结合的代价敏感学习算法

为了提取有效的代价敏感特征用于后续的分类器设计,研究者提出了代价敏感的特征选择算法.它的基本思路是:将错分代价信息嵌入特征选择过程,从而获得能够最小化错分损失的特征.

• CSLS 方法

CSLS(cost-sensitive Laplacian score)^[30,31]是一种应用于软件缺陷预测问题的代价敏感特征选择算法.CSLS 通过计算并排序每一个特征的代价敏感 Laplacian score 值,获得具有最小化错分损失能力的特征.具体地,采用如下公式计算第 t 个特征的代价敏感 Laplacian score 值:

$$L_t^{CSLS} = \sum_{i=1}^n \sum_{j=1}^n (-C_{l(x_i)l(x_j)}) (g_{ti} - g_{tj})^2 S_{ij} - \lambda \sum_{i=1}^n h(l(x_i)) (g_{ti} - m_t)^2 D_{ii} \quad (38)$$

其中, g_{ti} 表示第 i 个样本的第 t 个特征; S_{ij} 为无监督 K 近邻图的权重矩阵,其定义如公式(34)所示;重要性函数 $h(l(x_i))$ 的定义和公式(25)一致; m_t 为样本在第 t 个特征上的平均值; $D_{ii} = \sum_j S_{ij}$.

• DCSLS 方法

通过简单推导不难发现:CSLS 最大化了异类样本间的局部近邻结构,忽略了同类样本间的局部近邻结构.为了解决该问题,Wan 等人根据最小期望损失准则,提出了一种 DCSLS(discriminative cost-sensitive Laplacian score)^[85]方法,在保证异类样本间的代价敏感局部近邻关系最大化的同时,使得同类样本间的代价敏感局部近邻关系最小化.DCSLS 模型的定义如下:

$$L_t^{DCSLS} = \frac{(1 - \lambda) \sum_{i=1}^n h(l(x_i)) \sum_{j=1}^n (g_{ti} - g_{tj})^2 S_{ij}^W - \lambda \sum_{i=1}^n \sum_{j=1}^n C_{l(x_i)l(x_j)} (g_{ti} - g_{tj})^2 S_{ij}^B}{Var(g_t)} \quad (39)$$

其中, $Var(g_t)$ 为样本在第 t 个特征上的代价敏感方差; S_{ij}^W 表示同类样本的 K 近邻图的权重矩阵,其定义和公式(36)中的一致; S_{ij}^B 表示异类样本的 K 近邻图的权重矩阵,具体定义如下:

$$S_{ij}^B = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / \tau), & \text{if } x_i \text{ and } x_j \text{ are } K \text{ Nearest Neighbor and } l(x_i) \neq l(x_j) \\ 0, & \text{otherwise} \end{cases} \quad (40)$$

(4) 与半监督学习问题相结合的代价敏感学习算法

在上述的代价敏感学习算法中,研究者常假设样本的类别和代价矩阵已知,然后根据样本的类别信息分别给各样本赋予相应的代价.然而在实际应用中,获取样本的类别信息很难,需要付出大量的人力和物力.这里,代价敏感的半监督学习算法假设代价矩阵已知,研究只知道少量监督样本条件下的代价敏感学习问题.

- 代价敏感的半监督分类算法

- 1) CS⁴VM 方法

假设训练集中有 n_l 个监督样本和 n_u 个无监督样本.CS⁴VM(cost-sensitive semi-supervised support vector machine)^[86]通过定义如下的联合学习框架,用于交替学习代价敏感的分类器和无监督样本的标签信息:

$$\min_{f, \tilde{y}_i} : \frac{1}{2} \|f\|_2^2 + \gamma_1 \sum_{i=1}^{n_l} \text{loss}(f(x_i), y_i, C) + \gamma_2 \sum_{i=1}^{n_u} \text{loss}(f(x_i), \tilde{y}_i, C) \quad (41)$$

其中 f 是期望学得分类函数, \tilde{y}_i 为无监督样本的标签信息,代价敏感的损失函数 $\text{loss}(\cdot)$ 的定义和公式(19)一致, C 为代价矩阵.由于直接估计样本的标签信息可能会不准确,存在噪声以及错误标注问题,为此,CS⁴VM 通过采用估计无监督训练样本正负类均值的策略,解决无标记样本的标签预测问题.

- 2) CS-LapSVM 方法

为准确预测蛋白质中的 microRNA-binding Residues 问题,Zhou 等人提出了一种 CS-LapSVM(cost-sensitive Laplacian support vector machine)方法^[87].CS-LapSVM 通过在 CS-SVM 模型中嵌入保持样本局部流形结构的正则项,提高学习模型的判别能力.具体地,CS-LapSVM 模型的目标函数定义如下:

$$\min_f : \frac{1}{2} \|f\|_2^2 + \gamma_1 \sum_{i=1}^{n_l} \text{loss}(f(x_i), y_i, C) + \gamma_2 \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} \|f(x_i) - f(x_j)\|_2^2 S_{ij} \quad (42)$$

其中, S_{ij} 为无监督 K 近邻图的权重矩阵,具体定义如公式(34)所示.

- 3) SCS-LapSVM 方法

为了解决数据集中可能同时存在的代价敏感、类别不平衡、大量无标签样本以及噪声样本问题,SCS-LapSVM(sample-dependent cost-sensitive Laplacian support vector machine)^[88]首先采用无标签扩展策略预测无标签样本的标签信息,再将考虑了类别不平衡问题的错分代价嵌入到 Laplacian 支持向量机的经验损失和 Laplacian 正则项中.进一步,考虑到噪声样本对决策平面的影响,SCS-LapSVM 定义了一种样本依赖的代价,对噪声样本赋予较低的权重.

- 4) 代价敏感的标签扩展方法

SCS-LapSVM 采用两步学习的策略解决半监督代价敏感学习问题,然而其标签学习过程与代价敏感分类器的构建过程相互独立.预先学得的标签信息可能会影响后续的分类器构建,导致局部最优的分类结果.为解决该问题,Wan 等人提出了一种代价敏感的标签扩展(cost-sensitive label propagation,简称 CSLP)^[89]方法.CSLP 通过定义一种统一的代价敏感学习框架,用于联合学习代价敏感的分类模型 f 以及无监督训练样本的标签信息 Y_U .CSLP 方法的一般形式可定义如下:

$$\min_{f, Y} : \text{loss}(f, X, Y, C) \quad (43)$$

其中,训练样本的标签矩阵 $Y=[Y_L, Y_U]$ 由监督样本的标签矩阵 Y_L 和无监督样本的标签矩阵 Y_U 构成.

具体地,为了解决半监督代价敏感人脸识别问题,CSLP 首先采用矩阵分解技术获取人脸图像的高层语义表示,然后采用平方损失函数以及线性回归模型,定义满足最小错分损失准则的代价敏感隐语义回归模型;进一步,通过在 CSLP 目标函数中嵌入代价敏感的标签扩展正则化项,可获得鲁棒的样本标签信息.上述代价敏感的隐语义回归过程以及代价敏感的标签扩展过程交替执行,直至收敛,可获得最优的人脸识别结果.

- 代价敏感的半监督降维算法

- 1) CS³DA 方法

CS³DA(cost-sensitive semi-supervised discriminant analysis)^[90]被认为是最早提出的应用于代价敏感人脸识别问题的代价敏感半监督学习算法.为了有效利用训练集中的大量无监督样本,CS³DA 提出了两步学习的策略.

➤ 首先,基于 n_l 个监督样本 X_L ,稀疏重构无监督样本 X_U ,获得其稀疏表示系数 α :

$$\min_{\alpha_i} \|\alpha_i\|_1, \text{ s.t. } \|x_i - X_L \alpha_i\|_2^2 \leq \varepsilon \quad (44)$$

其中, $i=n_l+1, \dots, n$. ε 为稀疏表示的误差系数.

- 然后,利用系数 α_i 具有的判别性,预测无监督样本 x_i 的软标签信息 $y_i=[y_{1i}, y_{2i}, \dots, y_{ci}]^T$. 其中, y_{ki} 表示样本 x_i 属于第 k 类的概率.

当通过上述步骤获得所有无监督样本的标签信息后,CS³DA 采用和 CSLDA 相同的降维学习模型,学习代价敏感的投影方向 W .

2) PCSDA 方法

CS³DA 采用稀疏学习方法预测无监督样本的标签信息,其内在要求是人脸图像满足稀疏假设.然而,有研究者发现,人脸可能不是一个稀疏学习问题^[102].此外,CS³DA 没有从理论上证明其模型是否满足最小期望损失准则,这导致其求得的投影方向可能不具有控制错分损失的能力.为此,PCSDA(pairwise costs in semi-supervised discriminant analysis)^[91]首先采用比稀疏表示更为有效的 L_2 范数方法,预测无监督人脸图像的标签信息.具体地, L_2 范数方法可表述如下:

$$\min_{\beta_i} \|x_i - X_L \beta_i\|_2^2 \quad (45)$$

其中, β_i 为无监督样本 x_i 在 L_2 方法下的重构系数.值得注意的是:这里只需对监督样本执行一次 QR 分解 $X_L=QR$,即可同时求得所有无监督样本 X_U 的重构系数 $\beta=[\beta_{n_l+1}, \beta_{n_l+2}, \dots, \beta_n]=R^{-1}Q^T X_U$.然后,根据残差最小化准则获得无监督样本的标签信息.

进一步,PCSDA 通过定义满足成对贝叶斯风险准则的目标函数,使得学得的投影方向代价敏感.具体的目标函数定义如下:

$$\min_V \sum_{i=1}^{c-1} \sum_{j=i+1}^c p_i p_j \omega(\Delta_{ij}) \text{tr}(V^T S_W^{-1/2} (m_i - m_j)(m_i - m_j)^T S_W^{-1/2} V) \quad (46)$$

其中,代价敏感的类内离散度矩阵:

$$S_W = \sum_i^c h(i) S_{W_i}, \omega(\Delta_{ij}) = (P(ij) + P(ji)) \left(\frac{1}{2} + \frac{1}{2\Delta_{ij}^2} \text{erf} \left(\frac{\Delta_{ij}}{2\sqrt{2}} \right) \right), \Delta_{ij} = \sqrt{(m_i - m_j)^T (m_i - m_j)}, P(ij) = C_{ij} p_i, P(ji) = C_{ji} p_j.$$

p_i 和 m_i 分别为第 i 类的类先验概率和类均值.最终的投影方向为 $W = S_W^{-1/2} V$.

值得注意的是:如果令 $\omega(\Delta_{ij})=C_{ij}+C_{ji}$,PCSDA 模型就退化为 CS³DA;如果令 $\omega(\Delta_{ij}) = \frac{1}{2\Delta_{ij}^2} \text{erf} \left(\frac{\Delta_{ij}}{2\sqrt{2}} \right), h(i)=1$,

PCSDA 模型就退化为 aPAC(approximate pairwise accuracy criterion)^[103];如果令 $\omega(\Delta_{ij})=1, h(i)=1$,PCSDA 模型就退化为 LDA(linear discriminant analysis)^[40].

3) CS³CCA 方法

CS³CCA(cost-sensitive semi-supervised canonical correlation analysis)^[92]是一种应用于多视图学习问题的代价敏感半监督学习算法.CS³CCA 认为:如果在数据的某一视图中存在代价敏感问题,那么该问题也必将存在于其他视图中.例如在门禁系统中,无论是人脸视图还是在声音、指纹等其他视图中,都存在代价敏感学习问题.

这里,考虑到训练集中存在的大量无监督样本,CS³CCA 首先定义了一种基于 L_2 范数方法的软标签预测方法.和 PCSDA 一样,CS³CCA 可基于公式(45)获得无监督样本 x_i 的重构系数 β_i ,从而可得样本在第 k 类上的重构残差 $res_i(k) = \|x_i - X_L \delta_k(\beta_i)\|_2^2, k=1, 2, \dots, c$.进一步,通过引入如下的指数函数,估计无监督样本 x_i 的软标签信息 $[p(1|x_i), p(2|x_i), \dots, p(c|x_i)]^T$:

$$p(k | x_i) = \begin{cases} \lambda \exp(-\lambda \cdot res_i(k)), & \text{if } res_i(k) \geq 0 \\ 0, & \text{if } res_i(k) < 0 \end{cases} \quad (47)$$

其中, λ 为参数.类似地,重复上述过程,可得另一视图中无监督样本 y_i 的软标签信息 $[p(1|y_i), p(2|y_i), \dots, p(c|y_i)]^T$.

根据公式(47)所获得的软标签信息,CS³CCA 定义代价敏感的类间相关性矩阵 S_B^{cc} 和代价敏感的类内相关性矩阵 S_W^{cc} .具体定义如下:

$$\begin{cases} S_B^{cc} = \sum_{k=1}^c \sum_{k'=1, k' \neq k}^c C_{ij} \sum_{q=1}^n \sum_{q'=1}^n p(k | x_q) p(k' | y_{q'}) x_q y_{q'}^T \\ S_W^{cc} = \sum_{k=1}^c h(k) \sum_{q=1}^n \sum_{q'=1}^n p(k | x_q) p(k | y_{q'}) x_q y_{q'}^T \end{cases} \quad (48)$$

进一步,将 S_B^{cc} 和 S_W^{cc} 嵌入典型相关分析模型,可得 CS^3CCA 的目标函数:

$$\min_{W_x, W_y} : \frac{tr(W_x^T M_{xy} W_y)}{\sqrt{tr(W_x^T M_{xx} W_x)} \sqrt{tr(W_y^T M_{yy} W_y)}} \quad (49)$$

其中, $M_{xy} = S_W^{cc} - \alpha S_B^{cc}$, M_{xx} 和 M_{yy} 分别为各自视图上的离散度矩阵.

4 未来的研究趋势和待研究的难点问题

通过总结现有代价敏感学习的相关研究工作,未来可从如下几个方面开展研究.

(1) 研究困难条件下的代价敏感学习问题

现有的代价敏感的半监督学习算法^[86-92]假设代价矩阵已知,训练集中存在大量无监督样本.它们常采用两步学习的策略,即:先预测无标签训练样本的标签信息;然后根据代价矩阵定义无标签训练样本的错分代价,进而可定义监督的代价敏感学习算法.在两步学习过程中,标签预测过程代价不敏感且与后续的代价敏感学习模型相互独立.这使得学得标签可能会不利于后续的模式构建,导致局部最优的分类结果.因此在未来的研究中,应关注如何构建统一的代价敏感半监督学习模型.

进一步,在实际应用中,要获取样本的标签信息难,而要准确定义应用问题的错分代价矩阵则更难.现有的代价敏感学习算法常假设代价信息已知,由领域专家根据经验设定.但考虑到应用问题不断增加的复杂性,以及专家的经验知识可能存在局限性等因素,人工设定的代价可能不准确.针对该问题,有研究者将代价当作模型参数,采用交叉验证的方法加以学习^[3].但交叉验证的方法只能从预设的候选范围中选择代价,会导致局部最优解;而且交叉验证方法的时间复杂度较高.特别是在多分类问题中,由于错分类型较多,导致需要学习过多的代价参数.此外,还有学者将代价学习和支持向量机相结合,提出了代价区间的方法^[104,105],但该方法仍需事先设定代价区间,且只能应用于两分类问题.文献^[60,61]采用进化学习的方法优化求解错分代价,但仍存在局部最优解问题.

综上,在标签和代价信息未知的条件下,如何设计有效的代价敏感学习算法,是未来的一个重要研究内容.

(2) 研究统一的代价敏感学习模型

在直接的代价敏感学习算法中,研究者根据模型学习不同阶段,将代价信息分别嵌入到降维、特征选择以及分类器中,提出了代价敏感降维、代价敏感特征选择以及代价敏感分类算法.明显地,特征学习过程和分类器的构建过程相互独立.学得特征可能会影响后续的分类器设计,导致局部最优的分类结果.因此,如何将特征学习和分类器设计嵌入统一的代价敏感学习模型,是值得未来研究的一个重要内容.

进一步,如第 1.1 节代价的定义所述,在代价敏感学习中存在多种类型的代价,例如错分代价、计算代价以及存储代价,分别反映了学习模型的精度、时间复杂度以及空间复杂度这 3 个方面.在现有的机器学习算法中,研究者往往只考虑其中的一种代价,而忽略了另外两种代价.譬如在深度学习算法中,研究者往往强调其模型具有的高识别能力,对时空复杂度的关注度则不够.而是否具有较低的时空复杂度,往往是学习模型应用于实际时首要考虑的因素.在未来的工作中,是否能够将这 3 种代价统一考虑,设计一种综合考量学习算法精度、时间复杂度、空间复杂度的代价敏感模型?

(3) 研究面向多分类问题的代价敏感评价准则

为了评价学习模型的性能,代价敏感学习常以分类结果的总体错分损失作为评价准则.进一步,对于两分类问题,还可借鉴 ROC(receiver operating characteristic)曲线^[40]的思想绘制代价曲线^[13],用以准确度量各类的代价敏感错误率.然而在多分类问题中,绘制代价曲线较为困难.目前,研究者常通过定义多种评价准则,试图从多个不同的层面准确评价代价敏感的分类结果.例如,代价敏感人脸识别^[3,81-85,88-92]定义了如下 5 种评价准则,用以评价模型的整体错分损失、错误接收率、错误拒绝率、错误鉴别率以及整体的错误率:

$$\begin{cases} total\ cost = C_{IG} \times n_{fa} + C_{GI} \times n_{fr} + C_{GG} \times n_{fi} \\ Err_{IG} = n_{fa} / n_1^{te} \times 100\% \\ Err_{GI} = n_{fr} / n_G^{te} \times 100\% \\ Err_{GG} = n_{fi} / n_G^{te} \times 100\% \\ Err = (n_{fa} + n_{fr} + n_{fi}) / n_{te} \times 100\% \end{cases} \quad (50)$$

其中, n_{fa} 、 n_{fr} 、 n_{fi} 分别表示错分类型为错误接收、错误拒绝以及错误鉴别的样本数。 n_{te} 、 n_1^{te} 、 n_G^{te} 分别表示测试样本数以及其中的 gallery 和 imposter 类的样本数。

然而,评价指标的增加往往会带来决策上的困难,尤其是对非本领域的研究人员.在多分类问题中,是否可以设计一个直观、有效的评价指标,是未来值得研究的一个问题。

(5) 研究面向隐私保护问题的代价敏感学习模型

近年来,隐私保护问题引起了研究者的关注.隐私保护问题主要关注如何保护学习系统中的隐私信息不被恶意攻击者获取.例如,在基于生物特征的近邻搜索问题中^[106],研究者需要在保证检索精度不降低的情况下,保护生物特征数据库中的样本特征信息、样本间的距离信息以及样本的数据分布信息.因为这些信息如果泄露出去,攻击者就可以按照逆向工程的方法推测出整个数据库中的数据.这里,对于代价敏感学习算法,除了要保护数据库中样本的各类统计信息之外,还需要重点保护代价信息,因为其信息直接决定了分类器的分类超平面.如何在代价加密的条件下有效地优化求解学习模型,是一个难点问题。

5 总 结

代价敏感学习是机器学习领域的重要研究内容.针对不同的代价敏感应用问题,代价敏感学习可通过定义不同类型的代价,设计满足最小损失准则的学习模型.本文综述了代价敏感学习方法:首先,从代价的定义和代价敏感的风险决策这两方面介绍代价敏感学习的理论基础;然后介绍代价敏感学习的若干代表性应用领域,具体包括类别不平衡问题、人脸识别问题、多标记学习问题以及软件缺陷预测问题;进一步,根据机器学习模型训练的 3 个阶段,即训练数据输入、模型训练以及结果预测,详细介绍了 3 类代价敏感学习方法,即数据前处理方法、直接的代价敏感学习方法以及结果后处理方法;最后,对未来可能的研究趋势进行了展望。

References:

- [1] Elkan C. The foundations of cost-sensitive learning. In: Nebel B, ed. Proc. of the 17th Int'l Joint Conf. on Artificial Intelligence. Washington: Morgan Kaufmann Publishers, 2001. 973–978.
- [2] Ling XF, Sheng VS. A comparative study of cost-sensitive classifiers. Chinese Journal of Computers, 2007,30(8):1203–1212. [doi: 10.3321/j.issn:0254-4164.2007.08.001]
- [3] Zhang Y, Zhou ZH. Cost-sensitive face recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(10): 1758–1769. [doi: 10.1109/TPAMI.2009.195]
- [4] Lessmann S, Baesens B, Mues C, Pietsch S. Benchmarking classification models for software defect prediction: A proposed framework and novel findings. IEEE Trans. on Software Engineering, 2008,34(4):485–496. [doi: 10.1109/TSE.2008.35]
- [5] Turney PD. Types of cost in inductive concept learning. In: Langley P, ed. Proc. of the Workshop on Cost-sensitive Learning at the 17th Int'l Conf. on Machine Learning. Stanford: Morgan Kaufmann Publishers, 2000. 15–21.
- [6] Zhang SC. Cost-sensitive classification with respect to waiting cost. Knowledge Based Systems, 2010,23(5):369–378. [doi: 10.1016/j.knosys.2010.01.008]
- [7] Sheng VS. Fast data acquisition in cost-sensitive learning. In: Zaki MJ, Siebes A, Yu JX, Goethals B, Webb GI, Wu XD, eds. Proc. of the Int'l Conf. on Data Mining. Brussels: IEEE, 2012. 66–77. [doi: 10.1007/978-3-642-23184-1_6]
- [8] Sheng VS, Ling CX. Feature value acquisition in testing: A sequential batch test algorithm. In: Cohen WW, Moore A, eds. Proc. of the 23th Int'l Conf. on Machine Learning. Pittsburgh: ACM Press, 2006. 809–816. [doi: 10.1145/1143844.1143946]
- [9] Yang Q, Ling CX, Chai X, Pan R. Test-cost sensitive classification on data with missing values. IEEE Trans. on Knowledge and Data Engineering, 2006,18(5):626–638. [doi: 10.1109/TKDE.2006.84]

- [10] Ling CX, Yang Q, Wang J, Zhang SC. Decision trees with minimal costs. In: Brodley C, ed. Proc. of the Int'l Conf. on Machine Learning. Alberta: ACM Press, 2004. 69–76. [doi: 10.1145/1015330.1015369]
- [11] Chai X, Deng L, Yang Q, Ling CX. Test-cost sensitive Naïve Bayes classification. In: Bramer M, ed. Proc. of the Int'l Conf. on Data Mining. Brighton: IEEE, 2004. 51–58. [doi: 10.1109/ICDM.2004.10092]
- [12] Turney PD. Low size-complexity inductive logic programming: The east-west challenge considered as a problem in cost-sensitive classification. In: Muggleton S, Mizoguchi F, Furukawa K, eds. Proc. of the 5th Int'l Inductive Logic Programming Workshop. Tokyo: Springer-Verlag, 1995. 247–263.
- [13] Zhou ZH. Machine Learning. Beijing: Tsinghua University Press, 2016. 2–46 (in Chinese).
- [14] Margineantu D. On class probability estimates and cost-sensitive evaluation of classifiers. In: Langley P, ed. Proc. of the Workshop on Cost-Sensitive Learning at the 17th Int'l Conf. on Machine Learning. Stanford: Morgan Kaufmann Publishers, 2000. 15–21.
- [15] Zadrozny B, Elkan C. Learning and making decisions when costs and probabilities are both unknown. In: Lee D, Schkolnick M, Provost FJ, Srikant R, eds. Proc. of the 7th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM Press, 2001. 204–213. [doi: 10.1145/502512.502540]
- [16] Zadrozny B, Langford J, Abe N. A simple method for cost-sensitive learning. Technical Report, IBM, 2002.
- [17] Bahnsen AC, Aouada D, Ottersten B. Example-dependent cost-sensitive decision trees. Expert Systems with Applications, 2015, 42(19):6609–6619. [doi: 10.1016/j.eswa.2015.04.042]
- [18] Hastie T, Tibshirani P, Friedman J. The Elements of Statistical Learning. 2nd ed., New York: Springer-Verlag, 2001.
- [19] Wan JW. Cost-sensitive dimensionality reduction and its application in face recognition [Ph.D. Thesis]. Nanjing: Nanjing Normal University, 2013 (in Chinese with English abstract).
- [20] 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]
- [21] Lo HY, Wang JC, Lin SD. Cost-sensitive multi-label learning for audio tag annotation and retrieval. IEEE Trans. on Multimedia, 2011,13(3):518–529. [doi: 10.1109/tmm.2011.2129498]
- [22] Lo HY, Lin SD, Wang HM. Generalized k -Labelsets ensemble for multi-label and cost-sensitive classification. IEEE Trans. on Knowledge and Data Engineering, 2014,26(7):1679–1691. [doi: 10.1109/tkde.2013.112]
- [23] Fu ZL. Cost-sensitive ensemble learning algorithm for multi-label classification problems. Acta Automatica Sinica, 2014,40(6): 1075–1085 (in Chinese with English abstract). [doi: 10.3724/SP.J.1004.2014.01075]
- [24] Han L, Li M. Open source software classification using cost-sensitive multi-label learning. Ruan Jian Xue Bao/Journal of Software, 2014,25(9):1982–1991 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4639.htm> [doi: 10.13328/j.cnki.jos.004639]
- [25] Yuan J, Ming L, Zhou ZH. Software defect detection with ROCUS. Journal of Computer Science and Technology, 2011,26(2): 328–342. [doi: 10.1007/s11390-011-9439-0]
- [26] McCabe TJ. A complexity measure. IEEE Trans. on Software Engineering, 1976,2(4):308–320. [doi: 10.1109/TSE.1976.233837]
- [27] Halstead MH. Elements of Software Science. New York: Elsevier, 1977.
- [28] Chidamber SR, Kemerer CF. A metrics suite for object oriented design. IEEE Trans. on Software Engineering, 1994,20(6):476–493. [doi: 10.1109/32.295895]
- [29] Xing F, Guo P, Lyu R. A novel method for early software quality prediction based on support vector machine. In: Binder RV, Tsai JJP, eds. Proc. of the Int'l Symp. on Software Reliability Engineering. Chicago: IEEE, 2005. 213–222. [doi: 10.1109/ISSRE.2005.6]
- [30] Miao LS, Liu MX, Zhang DQ. Cost-sensitive feature selection with application in software defect prediction. In: Eklundh JO, Ohta YC, Tanimoto S, eds. Proc. of the Int'l Conf. on Pattern Recognition. Tsukuba: IEEE, 2012. 976–970.
- [31] Liu MX, Miao LS, Zhang DQ. Two-stage cost-sensitive learning for software defect prediction. IEEE Trans. on Reliability, 2014, 63(2):676–686. [doi: 10.1109/tr.2014.2316951]
- [32] Zheng J. Cost-sensitive boosting neural networks for software defect prediction. Expert Systems with Applications, 2010,37(6): 4537–4534. [doi: 10.1016/j.eswa.2009.12.056]
- [33] Arar OF, Ayan K. Software defect prediction using cost-sensitive neural network. Applied Soft Computing, 2015,33(C):263–277. [doi: 10.1016/j.asoc.2015.04.045]

- [34] Sheng VS, Ling CX. Thresholding for making classifiers cost-sensitive. In: Yolanda G, Raymond JM, eds. Proc. of the 21st AAAI Conf. on Artificial Intelligence. Boston: AAAI Press, 2006. 476–481.
- [35] Zadrozny B, Langford J, Abe N. Cost-sensitive learning by cost-proportionate example weighting. In: Shavlik J, ed. Proc. of the Int'l Conf. on Data Mining. Melbourne: IEEE, 2003. 435–442. [doi: 10.1109/ICDM.2003.1250950]
- [36] Seiffert C, Khoshgoftaar TM, Hulse JV, Napolitano A. A comparative study of data sampling and cost sensitive learning. In: Proc. of the Int'l Conf. on Data Mining. Pisa: IEEE, 2008. 46–52. [doi: 10.1109/ICDMW.2008.119]
- [37] Drummond C, Holte RC. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats oversampling. In: Fawcett T, Mishra N, eds. Proc. of the Workshop on Imbalanced Data Sets at the ICML 2003. Washington: AAAI Press, 2003. 1–8.
- [38] Zhou ZH, Liu XY. On multi-class cost-sensitive learning. Computational Intelligence, 2010,26(3):232–257. [doi: 10.1111/j.1467-8640.2010.00358.x]
- [39] Zhou ZH, Liu XY. On multi-class cost-sensitive learning. In: Yolanda G, Raymond JM, eds. Proc. of the 21th AAAI Conf. on Artificial Intelligence. Boston: AAAI Press, 2006. 567–572.
- [40] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed., New York: Wiley, 2001. [doi: 10.1007/978-1-4471-0409-4_3]
- [41] Ting KM. An instance-weighting method to induce cost-sensitive trees. IEEE Trans. on Knowledge and Data Engineering, 2002, 14(3):659–665. [doi: 10.1109/TKDE.2002.1000348]
- [42] Liu XY, Zhou ZH. The influence of class imbalance on cost-sensitive learning: An empirical study. In: Liu JM, Wah BW, eds. Proc. of the 6th Int'l Conf. on Data Mining. Hong Kong: IEEE, 2006. 970–974. [doi: 10.1109/ICDM.2006.158]
- [43] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligent Research, 2002,16:321–357. [doi: 10.1613/jair.953]
- [44] Wilson DL. Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans. on Systems, Man and Cybernetics, 1972, 2:408–421. [doi: 10.1109/tsmc.1972.4309137]
- [45] Lopez V, Fernandez A, Moreno-Torres JG, Herrera F. Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. Expert Systems with Applications, 2012,39(7):6585–6608. [doi: 10.1016/j.eswa.2011.12.043]
- [46] Domingos P. MetaCost: A general method for making classifiers cost-sensitive. In: Fayyad UM, Chaudhuri S, Madigan D, eds. Proc. of the Int'l Conf. on Knowledge Discovery and Data Mining. San Diego: ACM Press, 1999. 155–164. [doi: 10.1145/312129.312220]
- [47] Zhou ZH, Liu XY. Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Trans. on Knowledge and Data Engineering, 2006,18(1):63–77. [doi: 10.1109/tkde.2006.17]
- [48] Lomax S, Vadera S. A survey of cost-sensitive decision tree induction algorithms. ACM Computing Surveys, 2013,45(2):1–35. [doi: 10.1145/2431211.2431215]
- [49] Vadera S. CSNL: A cost-sensitive non-linear decision tree algorithm. ACM Trans. on Knowledge Discovery from Data. 2010,4(2): 1–35. [doi: 10.1145/1754428.1754429]
- [50] Freitas A, Costa-Pereira A, Brazdil P. Cost-sensitive decision trees applied to medical data. In: Song Y, Eder J, Nguyen TM, eds. Proc. of the Int'l Conf. on Data Warehousing and Knowledge Discovery. Regensburg: Springer-Verlag, 2007. 303–312. [doi: 10.1007/978-3-540-74553-2_28]
- [51] Davis JV, Jungwoo H, Rossbach CJ. Cost-sensitive decision tree learning for forensic classification. In: Furnkranz J, Scheffer T, Spiliopoulou M, ed. Proc. of the European Conf. on Machine Learning. Berlin: Springer-Verlag, 2006. 622–629. [doi: 10.1007/11871842_60]
- [52] Turney PD. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction. Journal of Artificial Intelligence Research, 1995,2:369–409. [doi: 10.1111/j.1365-2958.2009.06876.x]
- [53] Ling CX, Sheng VS, Yang Q. Test strategies for cost-sensitive decision trees. IEEE Trans. on Knowledge and Data Engineering, 2006,18(8):1055–1067. [doi: 10.1109/TKDE.2006.131]
- [54] Zhao SW, Zuo L, Wang SL, Shen LS. A multi-objective optimization based constructing cost-sensitive decision trees method. Acta Electronica Sinica, 2011,39(10):2348–2396 (in Chinese with English abstract).

- [55] Zhang SC, Zhu XF, Zhang JL, Zhang CQ. Cost-time sensitive decision tree with missing values. In: Zhang ZL, Siekmann JH, eds. Proc. of the Int'l Conf. on Knowledge Science, Engineering and Management. Melbourne: Springer-Verlag, 2009. 294–297. [doi: 10.1007/978-3-540-76719-0_44]
- [56] Wang T, Qin ZX, Jin Z, Zhang SC. Handling over-fitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning. The Journal of Systems and Software, 2010,83:1137–1147. [doi: 10.1016/j.jss.2010.01.002]
- [57] Morik K, Brochhausen P, Joachims T. Combining statistical learning with a knowledge-based approach: A case study in intensive care monitoring. In: Bratko I, Dzeroski S, eds. Proc. of the 16th Int'l Conf. on Machine Learning. Bled: Morgan Kaufmann Publishers, 1999. 268–277. [doi: 10.17877/DE290R-3088]
- [58] Masnadi-Shirazi H, Vasconcelos N, Iranmehr A. Cost-sensitive support vector machines. Neurocomputing, 2019. [doi: 10.1016/j.neucom.2018.11.099]
- [59] Lee Y, Lin Y, Wahba G. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. Journal of American Statistical Association, 2004,99(465):67–81. [doi: 10.2307/27590354]
- [60] Gu B, Sheng VS, Tay K, Romano W, Li S. Cross validation through two-dimensional solution surface for cost-sensitive SVM. IEEE Trans. on Pattern Analysis Machine Intelligence, 2017,39(6):1103–1121. [doi: 10.1109/TPAMI.2016.2578326]
- [61] Sun YM, Kamel MS, Wong AK, Wang Y. Cost-sensitive boosting for classification of imbalanced data. Pattern Recognition, 2007, 40(12):3358–3378. [doi: 10.1016/j.patcog.2007.04.009]
- [62] Zhang C, Tan C, Li HZ, Hong GS. A cost-sensitive deep belief network for imbalanced classification. IEEE Trans. on Neural Networks and Learning System, 2018,30(1):109–122. [doi: 10.1109/TNNLS.2018.2832648]
- [63] Zhang GQ, Sun HJ, Ji ZX, Yuan YH, Sun QS. Cost-sensitive dictionary learning for face recognition. Pattern Recognition, 2016,60: 613–629. [doi: 110.1016/j.patcog.2016.06.012]
- [64] Man JY, Jing XY, Zhang D, Lan C. Sparse cost-sensitive classifier with application to face recognition. In: Macq B, Schelkens P, eds. Proc. of the Int'l Conf. on Image Processing. Brussels: IEEE, 2011. 1773–1776. [doi: 10.1109/ICIP.2011.6115804]
- [65] Wan JW, Yang M, Wang HY. Cost sensitive matrix factorization for face recognition. In: Yin HJ, Gao Y, Chen SC, Wen YM, Cai GY, Gu TL, Du JP, Tallon-Ballesteros AJ, Zhang ML, eds. Proc. of the Int'l Conf. on Intelligent Data Engineering and Automated Learning. Guilin: Springer-Verlag, 2017. 136–145. [doi: 10.1007/978-3-319-68935-7_16]
- [66] Zheng J. Cost-sensitive boosting neural networks for software defect prediction. Expert Systems with Applications, 2010,37(6): 4537–4543. [doi: 10.1016/j.eswa.2009.12.056]
- [67] Ryu D, Jang J, Baik J. A transfer cost-sensitive boosting approach for cross-project defect prediction. Software Quality Journal, 2017,25(1):1–38. [doi: 10.1007/s11219-015-9287-1]
- [68] Li ZQ, Jing XY, Wu F, Zhu XK, Xu BW, Ying S. Cost-sensitive transfer kernel canonical correlation analysis for heterogeneous defect prediction. Automated Software Engineering, 2018,25(2):201–245. [doi: 10.1007/s10515-017-0220-7]
- [69] Herbold S. Benchmarking cross-project defect prediction approaches with costs metrics. ArXIV: 1801.04107V.
- [70] Li ZQ, Jing XY, Zhu XK. Heterogeneous fault prediction with cost-sensitive domain adaptation. Software Testing, Verification and Reliability, 2018,28(2):1–22. [doi: 10.1002/stvr.1658]
- [71] Wu F, Jing XY, Sun Y, Sun J, Huang L, Cui FY, Sun YF. Cross-project and within-project semisupervised software defect prediction: A unified approach. IEEE Trans. on Reliability, 2018,67(2):581–597. [doi: 10.1109/TR.2018.2804922]
- [72] Sahin Y, Bulkan S, Duman E. A cost-sensitive decision tree approach for fraud detection. Expert System with Application, 2013, 40(15):5916–5923. [doi: 10.1016/j.eswa.2013.05.021]
- [73] Zhou B, Yao YY, Luo JG. Cost-sensitive three-way email spam filtering. Journal of Intelligence Information System, 2014,42(1): 19–45. [doi: 10.1007/s10844-013-0254-7]
- [74] Artan Y, Haider MA, Langer DL, Kwast TH, Evans AJ, Yang Y, Wernick MN. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. IEEE Trans. on Image Processing, 2010,19(9): 2444–2455. [doi: 10.1109/TIP.2010.2048612]
- [75] Hinton G, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computing, 2006,18(7):1527–1554. [doi: 10.1162/neco.2006.18.7.1527]
- [76] Wright J, Yang A, Sastry S, Ma Y. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(2):210–227. [doi: 10.1109/TPAMI.2008.79]

- [77] Hu WM, Ding XM, Li B, Wang JC, Gao Y, Wang FS. Multi-perspective cost-sensitive context-aware multi-instance sparse coding and its application to sensitive video recognition. *IEEE Trans. on Multimedia*, 2015,18(1):76–89. [doi: 10.1109/TMM.2015.2496372]
- [78] Huang KH, Lin HT. Cost-sensitive label embedding for multi-label classification. *Machine Learning*, 2017,106(9-10):1725–1746. [doi: 10.1007/s10994-017-5659-z]
- [79] Yang YY, Huang KH, Chang CW, Lin HT. Cost-sensitive reference pair encoding for multi-label learning. In: Phung DQ, Tseng VS, Webb GI, Ho B, Ganji M, Rashidi L, eds. *Proc. of the Pacific-Asia Conf. on Knowledge Discovery & Data Mining*. Melbourne: Springer-Verlag, 2018. 143–145. [doi: 10.1007/978-3-319-93034-3_12]
- [80] Hsieh CY, Lin YA, Lin HT. A deep model with local surrogate loss for general cost-sensitive multi-label learning. In: McIlraith SA, Weinberger KQ, eds. *Proc. of the AAAI Conf. on Artificial Intelligence*. New Orleans: AAAI Press, 2018. 3239–3246.
- [81] Lu JW, Tan YP. Cost-sensitive subspace learning for face recognition. In: Darrell T, Hogg D, Jacobs D, eds. *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition*. San Francisco: IEEE, 2010. 2661–2666. [doi: 10.1109/CVPR.2010.5539983]
- [82] Lu JW, Tan YP. Cost-sensitive subspace learning for human age estimation. In: Siu WC, ed. *Proc. of the 17th Int'l Conf. on Image Processing*. Hong Kong: IEEE, 2010. 1593–1596. [doi: 10.1109/ICIP.2010.5650873]
- [83] Wan JW, Yang M, Ji GL, Chen YJ. Weighted cost sensitive locality preserving projection for face recognition. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(5):1155–1164 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4263.htm> [doi: 10.3724/sp.j.1001.2013.04263]
- [84] Wan JW, Yang M. Pairwise costs in subclass discriminant analysis. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(11): 2597–2609 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4473.htm> [doi: 10.3724/SP.J.1001.2013.04473]
- [85] Wan JW, Yang M, Chen YJ. Discriminative cost sensitive Laplacian score for face recognition. *Neurocomputing*, 2015,152: 333–344. [doi: 10.1016/j.neucom.2014.10.059]
- [86] Li YF, Kwok J, Zhou ZH. Cost-sensitive semi-supervised support vector machine. In: Fox M, Poole D, eds. *Proc. of the 24th AAAI Conf. on Artificial Intelligence*. Atlanta: AAAI Press, 2010. 500–505.
- [87] Wu JS, Zhou ZH. Sequence-based prediction of microRNA-binding residues in proteins using cost-sensitive laplacian support vector machines. *ACM/IEEE Trans. on Computational Biology and Bioinformatics*, 2013,10(3):752–759. [doi: 10.1109/tcbb.2013.75]
- [88] Wan JW, Yang M, Ji GL, Chen YJ. Cost sensitive semi-supervised laplacian support vector machine. *Acta Electronica Sinica*, 2012,40(7):1410–1415 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2012.07.020]
- [89] Wan JW, Wang Y. Cost-sensitive label propagation for semi-supervised face recognition. *IEEE Trans. on Information Forensics and Security*, 2019,14(7):1729–1743. [doi: 10.1109/TIFS.2018.2885252]
- [90] Lu JW, Zhou XZ, Tan YP. Cost-sensitive semi-supervised discriminant analysis for face recognition. *IEEE Trans. on Information Forensics and Security*, 2012,7(3):944–953. [doi: 10.1109/tifs.2012.2188389]
- [91] Wan JW, Yang M, Gao Y, Chen YJ. Pairwise costs in semi-supervised linear discriminant analysis for face recognition. *IEEE Trans. on Information Forensics and Security*, 2014,9(10):1569–1580. [doi: 10.1109/tifs.2014.2343833]
- [92] Wan JW, Wang HY, Yang M. Cost sensitive semi-supervised canonical correlation analysis for multi-view dimensionality reduction. *Neural Processing Letters*, 2017,45(2):411–430. [doi: 10.1007/s11063-016-9532-z]
- [93] Huang SJ, Zhao JW, Liu ZY. Cost-effective training of deep CNNs with active model adaptation. In: Guo YK, Farooq F, eds. *Proc. of the 24th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining*. London: ACM Press, 2018. 1580–1588. [doi: 10.1145/3219819.3220026]
- [94] Yan YF, Huang SJ. Cost-effective active learning for hierarchical multi-label classification. In: Lang J, ed. *Proc. of the Int'l Joint Conf. on Artificial Intelligence*. Stockholm: Morgan Kaufmann Publishers, 2018. 2962–2968. [doi: 10.24963/ijcai.2018/411]
- [95] Huang SJ, Chen JL, Mu X, Zhou ZH. Cost-effective active learning from diverse labelers. In: Sierra C, ed. *Proc. of the Int'l Joint Conf. on Artificial Intelligence*. Melbourne: Morgan Kaufmann Publishers, 2017. 1879–1885. [doi: 10.24963/ijcai.2017/261]
- [96] Wang HS, Cui ZC, Chen YX, Avidan M, Abdallah AB, Kronzer A. Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, 2018,15(6):1986–1978. [doi: 10.1109/TCBB.2018.2827029]
- [97] Zhang GQ, Sun HJ, Xia GY, Sun QS. Multiple kernel sparse representation-based orthogonal discriminative projection and its cost-sensitive extension. *IEEE Trans. on Image Processing*, 2016,25(9):4271–4285. [doi: 10.1109/TIP.2016.2587119]

- [98] Tsoumakas G, Katakis I, Vlahavas I. Random k -labelsets for multilabel classification. *IEEE Trans. on Knowledge and Data Engineering*, 2011,23:1079–1089. [doi: 10.1109/TKDE.2010.164]
- [99] Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data. In: Maimon O, Rokach L, eds. *Data Mining and Knowledge Discovery Handbook*. Boston: Springer-Verlag, 2010. [doi: 10.1007/978-0-387-09823-4_34]
- [100] He XF, Niyogi P. Locality preserving projections. In: Thrun S, Saul LK, Scholkopf B, eds. *Proc. of the Neural Information Processing Systems*. British Columbia: MIT Press, 2003. 153–160. [doi: 10.1016/j.patcog.2011.05.014]
- [101] Zhu M, Martinez AM. Subclass discriminant analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006,28(8): 1274–1286. [doi: 10.1109/TPAMI.2006.172]
- [102] Shi QF, Eriksson A, Shen CH. Is face recognition really a compressive sensing problem? In: *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition*. Colorado Springs: IEEE, 2011. 553–560. [doi: 10.1109/CVPR.2011.5995556]
- [103] Loog M, Duin RP, Umbach RH. Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(7):762–766. [doi: 10.1109/34.935849]
- [104] Zhou YH, Zhou ZH. Large margin distribution learning with cost interval and unlabeled data. *IEEE Trans. on Knowledge and Data Engineering*, 2016,28(7):1749–1763. [doi: 10.1109/TKDE.2016.2535283]
- [105] Liu XY, Zhou ZH. Learning with cost intervals. In: Rao B, Krishnapuram B, Tomkins A, Yang Q, eds. *Proc. of the 16th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining*. Washington: ACM Press, 2010. 403–412. [doi: 10.1145/1835804.1835857]
- [106] Wang Y, Wan JW, Guo J, Cheung YM, Yuen PC. Inference-based similarity search in randomized montgomery domains for privacy-preserving biometric identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018,40(7):1611–1624. [doi: 10.1109/tpami.2017.2727048]

附中文参考文献:

- [13] 周志华.机器学习.北京:清华大学出版社,2016.2–46.
- [19] 万建武.代价敏感降维及其人脸识别应用研究[博士学位论文].南京:南京师范大学,2013.
- [23] 付忠良.多标签代价敏感分类集成学习算法.自动化学报,2014,40(6):1075–1085. [doi: 10.3724/SP.J.1004.2014.01075]
- [24] 韩乐,黎铭.基于代价敏感多标签学习的开源软件分类.软件学报,2014,25(9):1982–1991. <http://www.jos.org.cn/1000-9825/4639.htm> [doi: 10.13328/j.cnki.jos.004639]
- [54] 赵士伟,卓力,王素玉,沈兰荪.一种基于 NNIA 多目标优化的代价敏感决策树构建方法.电子学报,2011,39(10):2348–2396.
- [83] 万建武,杨明,吉根林,陈银娟.一种面向人脸识别的加权代价敏感局部保持投影.软件学报,2013,24(5):1155–1164. <http://www.jos.org.cn/1000-9825/4263.htm> [doi: 10.3724/sp.j.1001.2013.04263]
- [84] 万建武,杨明.一种引入成对代价的子类判别分析.软件学报,2013,24(11):2597–2609. <http://www.jos.org.cn/1000-9825/4473.htm> [doi: 10.3724/SP.J.1001.2013.04473]
- [88] 万建武,杨明,吉根林,陈银娟.代价敏感的半监督 Laplacian 支持向量机.电子学报,2012,40(7):1410–1415. [doi: 10.3969/j.issn.0372-2112.2012.07.020]



万建武(1986—),男,江苏常州人,博士,讲师,CCF 专业会员,主要研究领域为数据挖掘,机器学习,模式识别.



杨明(1964—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为数据挖掘,机器学习,模式识别.