第50卷第7期 中国科学技术大学学报 Vol. 50, No. 7
2020年7月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Jul. 2020
文章编号:0253-2778(2020)07-0968-09

不均衡数据情形的基于聚焦损失的CGAN的集成分类方法

崔文泉,余厚莹,侯晓天

(中国科学技术大学管理学院统计与金融系,安徽合肥 230026)

摘要:针对非均衡数据的情形,基于条件生成对抗网络(conditional generative adversarial networks,CGAN),利用梯度提升树研究了聚焦损失的CGAN的集成分类方法.该方法首先通过CGAN降低不均衡率,通过聚焦损失的权值均衡结合GBDT算法,适当增加对少数类样本的关注度进而进一步提升分类器的分类性能.对方法的性质进行了研究,获得了若干理论成果.证明了:在一定条件下,由CGAN产生的经验条件分布收敛于相应总体的条件分布;聚集损失的CGAN方法其经验风险收敛到期望风险;该方法的估计量会收敛到使得期望风险最小化的函数.实验结果显示了聚焦损失的CGAN方法具有良好的表现.

关键词:非均衡数据;条件生成对抗网络;聚焦损失;集成学习
中图分类号:TP181;O212.1 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2020.07.014
2010 Mathematics Subject Classification:62H30

引用格式:崔文泉,余厚莹,侯晓天. 不均衡数据情形的基于聚焦损失的CGAN的集成分类方法[J]. 中国科学技术大学学报,2020,50(7):968-976.
CUI Wenquan, YU Houying, HOU Xiaotian. Focused loss-based for imbalanced data scenarios integrated classification methods for CGAN[J]. Journal of University of Science and Technology of China, 2020,50(7):968-976.

Focused loss-based for imbalanced data scenarios integrated classification methods for CGAN

CUI Wenquan, YU Houying, HOU Xiaotian

(*Department of Statistics and Finance, School of Management, University of Science and of Technology of China, Hefei* 230026, *China*)

Abstract: For the case of imbalanced data, an integrated classification method for CGAN-focal-loss was investigated based on conditional generative adversarial networks (CGAN) using gradient boosting trees. The method first reduces the imbalance rate by CGAN, and further improves the classification performance of the classifier by increasing the focus on a few classes of samples through the weight balancing of the focused loss combined with the GBDT algorithm. The properties of the method were investigated and several theoretical results were obtained. It was proved that the empirical conditional distribution generated by CGAN converges to the conditional distribution of the corresponding aggregate under certain conditions; that the empirical risk of the CGAN method with focused loss converges to the expected risk; and that the estimator of the method converges to the function that minimizes the expected risk. The experimental results show the good performance of the CGAN-focal-loss method.

Key words: imbalanced data; conditional generative adversarial networks (CGAN); focal loss; ensemble learning

0 引言

解决类别不均衡的分类问题是学术界和业界一个重要的研究方向[1].类别不均衡问题出现在现实生活中的方方面面,比如医疗诊断[2]、人脸识别[3]、邮箱信件分类[4]、金融欺诈检测[5]和半导体工艺[6]等.对不均衡数据的分类问题,若不进行有针对性处理,通常构建的分类器对新个体的分类倾向于判为多数类,也即对少数类样本难以识别[7].然而,在实际问题中,对少数类个体的甄别往往是更为重要的.因此,需要改进经典的分类方法以提高对少数类个体的分类性能.

近年来,学者们提出了许多解决类别不均衡问题的方法,这些方法主要分为三个层面[8]:数据、算法以及关于数据和算法结合的层面.其中数据层面主要通过欠抽样、过抽样等抽样的方法来解决类别

收稿日期:2020-05-25;**修回日期**:2020-06-27
基金项目:国家自然科学基金(71873128)资助.
作者简介:崔文泉(通讯作者),男,1964年生,博士/副教授.研究方向:数理统计.E-mail:wqcui@ustc.edu.cn

不均衡的问题[9]. 欠抽样主要是随机欠抽样(random undersampling,RUS),通过对多数类样本进行随机删减,使得数据趋于平衡. 但该方法随机地舍弃多数类样本中潜在的有用信息而使得模型分类性能不高. 对于少数类样本极少的情况下,该类方法会失效. 于是,出现了通过增加少数类样本使得数据均衡的过抽样方法. 典型的过抽样方法包括随机过抽样(random oversampling, ROS)、SMOTE (synthetic minority oversampling technique)[10]、borderline-SMOTE[11]以及自适应样本合成方法(adaptive synthetic sampling approach, ADASYN)[12]. 随机过抽样通过对少数类样本随机抽样来补充不足类别的样本. SMOTE 算法的基本思想是对每个少数类样本点和它最近邻的连线上随机选一点作为新的少数类样本点,使得数据趋于均衡[13]. Borderline-SMOTE 是在 SMOTE 基础上改进的过抽样算法,该算法是使用边界上少数类样本来合成新样本[11]. ADASYN 也是 SMOTE 的一个扩展,它是通过 k 近邻来生成少数类样本的人工数据的方法[12]. 其他抽样方法被主要包括:SMOTETomek 和 SMOTEENN 算法[13],SMOTETomek 和 SMOTEENN 算法均是在 SMOTE 的基础上进行的. SMOTETomek 最后剔除数据集中边界处的噪音样本,使分类界面更清晰[13]. SMOTEENN 算法则对新数据集中的每一个样本点使用 k 最近邻得到每个样本的预测结果,若预测类别和实际类别不符,则剔除该样本. 第二个层面是基于算法层面的研究. 该类方法结合不均衡数据的特点,对现有的分类算法进行适当的改进,使其在训练过程中提高对少数类样本的识别率. 代价敏感学习算法(cost-sensitive learning)是在算法层面解决类别不均衡问题常用的手段,通过引入代价来指导学习的过程. 代价敏感学习算法的核心思想是要找到一个合适的代价矩阵[14], 通过代价矩阵给少数类样本一个较高的错分代价因子,以全局误分率最低为优化目标来提高分类结果. 最后一个层面将数据层面和算法层面进行整合,以提高分类算法的分类效果[15].

本文关注的是从数据和算法结合的层面来解决类别不均衡问题. 最近,基于 GANs 提出了学习少数类样本来生成人工数据的方法[8,16,22,23,25,27]. 与传统的抽样方法相比,GANs 通过生成器来生成人工数据,其目的在于生成与少数类样本相似的数据,混淆判别器使其难以判别. GANs 可以生成逼近真实数据分布的样本,将其补充到少数类数据集中. GANs 中的方法有 BAGAN (balancing GAN)[16]、DCGAN[17]等作为一种数据增强的工具来均衡数据集,以便提高分类的准确性. 随着研究的深入,已有学者通过 VGrow(variational gradient flow)算法[18]说明了由向量场控制的分布会渐进收敛于目标分布,也可以用来作为数据增强的工具. 并且深度学习框架包含的 GAN 方法还可以用于处理非均衡的多分类问题[19]. 同时,在机器学习领域中,对损失函数的改进也是一个引人关注的问题. Lin 等[20]通过聚焦损失(focal loss)来解决分类问题中类别不均衡以及判别样本难易程度的问题. ROC (receiver operating characteristic)曲线是对不均衡数据集分类性能评价的综合指标[21]. ROC 曲线下方的面积为 AUC(area under the ROC),用标量值 AUC 来量化 ROC. 因此 AUC 是一种从总体上评价分类器性能的便利指标,可较为客观地评价分类器的效果.

在数据层面,对具有非均衡特性的分类问题,常见的方法主要是基于重抽样技术. 其方法是对数据的简单复制、剔除或从样本点的局部出发去合成样本,进而改变了原数据集的结构,从而带来一定偏差. 而 CGAN 是从数据集的全局出发,能够利用数据集的全部信息生成逼近真实数据分布的样本. 因此,已有的文献通过 GANs 相关方法[8,22-23]来处理金融数据类别不均衡的问题. 虽然分类效果有一定的提升,但仍有不足之处. 首先大多数 GANs 方法只是从实验的角度说明方法的可行性,对所提的方法缺乏理论结果. 其次对于不同类别的不均衡数据只是通过 GANs 的相关方法使数据均衡,可能会带来一定的偏差. 在算法层面,常用的方法就是代价敏感学习算法,即在算法迭代过程中,通过对少数类样本被错分时给予较高的代价损失. 可是,寻找合适的代价矩阵比较困难,普遍适用性较弱.

基于以上考虑,为了提高模型的普遍适用性,我们对聚焦损失的 CGAN 进行方法研究,来说明该方法可以处理不均衡数据的问题. 首先使用 CGAN 生成少数类样本,缩小不均衡比率. 之后通过 GBDT 集成学习框架[24],使用基分类器是二分类的决策树,利用聚焦损失权值均衡的思想,使得模型在训练过程中不仅能控制正负样本的权重,还能控制难易分类的样本权重以提高算法的分类能力.

1　方法及性质

1.1　CGAN 方法

GANs 是一类通用、灵活、强大的生成式深度学习模型构成的框架,是一种十分高效的生成数据的方法. 生成对抗网络(GAN)是由 Goodfellow 等[25]在 2014 年提出来的,其模型主要包括两个基本结构——判别器 (discriminator) 和生成器 (generator). 其本质是生成器 G 和判别器 D 在博弈过程中不断优化直至实现纳什均衡[26]. 也就是判别器 D 无法区分观测数据和生成器 G 生成的数据,此时的生成器 G 已经学习到观测数据的分布特点.

Mirza 等[27]提出了条件生成对抗网络(CGAN),CGAN 在 GAN 的基础上利用了响应变量的信息,相当于把无监督的 GAN 变成有监督的 CGAN 模型. 也即在由随机产生的数据通过 GAN 的方式得到与输入数据难以区分的新数据的过程中,CGAN 利用了响应变量与输入变量之间关系的信息,这样使得数据产生更加具有针对性,比没有考虑响应变量的 GAN 方法具有更高的效率. CGAN 相对 GAN 引入了一个空间 $\mathcal{Y}$, 表示来自训练数据的条件信息. 记 $G:\mathcal{Z}\times\mathcal{Y}\to\mathcal{X}$ 的生成模型,其

中 $\mathcal{Z}$ 为生成器 G 输入的随机噪音 Z 的取值空间，对任意给定的 Y，将 Z 映射到观测数据的样本空间 $\mathcal{X}$，目的是为了捕获训练数据总体的条件分布. $D:\mathcal{X}\times\mathcal{Y}\to[0,1]$ 的判别模型，估计输入样本来自训练数据的概率. 其中 P_Z 表示随机噪声的分布，P_X 表示观测数据的分布，E 表示期望，目标函数 $V(D,G)$ 为

$$V(D,G)=E_{X\sim p_X}[\log D(X\mid Y)]+E_{Z\sim p_Z}[\log(1-D(G(Z\mid Y)))] \tag{1}$$

对此目标函数进行极小极大的优化处理. 也即，先对 D 进行极大化处理，再对 G 进行极小化处理，最终得到理论上的最优解.

1.2 聚焦损失的 CGAN 学习方法

本文主要研究具有非均衡特性的二分类问题. 通过分析数据的不均衡率，基于 CGAN、聚焦损失以及 GBDT 算法来探究聚焦损失的 CGAN 方法.

设 $(X,Y)\in\mathbb{R}^d\times\{1,0\}$，记其训练样本为 $\Delta=\{(x_i,y_i)\}_{i=1}^n$，且 $\sum_{i=1}^n 1_{\{y_i=0\}}\gg\sum_{i=1}^n 1_{\{y_i=1\}}$. $\mathcal{G}=\{G_\theta:\theta\in\Theta\subset\mathbb{R}^p\}$ 是一族生成器，$\mathcal{D}=\{D_\alpha:\alpha\in\Lambda\subset\mathbb{R}^q\}$ 是一族判别器. 鉴于多层感知机在处理数据时具有良好的表现[8,22]，本文中 CGAN 在训练过程中 $\mathcal{G}$ 和 $\mathcal{D}$ 选用多层感知机网络结构，比较适合进行非线性拟合函数. 记 l 层的输入层为 $a^{[l-1]}$，输出层为 $a^{[l]}$，权重记为 $W^{[l]}$，偏置记为 $b^{[l]}$，该层的神经元个数为 $n^{[l]}$，激活函数记为 σ. 则层与层之间的关系为 $a^{[l]}=\sigma(W^{[l]}a^{[l-1]}+b^{[l]})$，且输出、权重以及偏置的维度分别为 $(n^{[l]},1)$，$(n^{[l]},n^{[l-1]})$，$(n^{[l]},1)$. 也即训练生成器 G 和判别器 D 就是估计每层参数 W,b 的过程，将每层所有参数放在一起分别对应记为 θ 和 α. 其中生成器 G 和判别器 D 的隐藏层的激活函数 σ 采用 tanh 激活函数. 模型选用多层感知机结构，层与层之间均是通过全连接的方式进行连接的. 对于含有大量参数的多层感知机网络过拟合是一个非常严重的问题. 基于这一问题，Dropout 算法[28]是神经网络中常用的正则化方法，适合含有大量参数的模型. Dropout 是指神经网络在进行学习时，在每次迭代训练过程中，每一层按照一定的概率值对网络中的神经元进行随机消除，也即在每一次训练过程中通过限制神经元的个数来简化复杂网络模型，从而能有效防止过拟合.

由式(1)，CGAN 目标函数的经验形式为

$$\hat{V}(D_\alpha,G_\theta)=\frac{1}{n}\sum_{i=1}^n\log D_\alpha(X_i\mid Y_i)+\frac{1}{n}\sum_{i=1}^n\log(1-D_\alpha(G_\theta(Z_i\mid Y_i))).$$

Friedman[36]提出了梯度提升算法来拟合基分类器. 算法层面本文使用决策树为基分类器，把 GBDT 集成算法的损失函数取为聚焦损失[20]，表达式为

$$L(y,f(x))=\begin{cases}-\alpha(1-f(x))^\gamma\log f(x), & y=1;\\ -(1-\alpha)f(x)^\gamma\log(1-f(x)), & y=0\end{cases} \tag{2}$$

式中，$f(x)$ 是模型在给定 x 条件下预测 $y=1$ 的概率. 聚焦损失在经典交叉熵的基础上对类别 1 添加了参数 α 以及动态缩放因子 γ. 不对数据的非均衡性进行有针对性处理时，通常构建的分类器对新个体的分类趋向于判为多数类. 对于类别不均衡数据进行有针对处理时，聚焦损失一方面通过设定参数 $\alpha\in(0,1)$ 来调节少数类样本的权重，通过增加 α 来提高少数类样本对总体损失的贡献. 另一方面，聚焦损失通过 $\gamma\geqslant 0$ 这个可调的超参数控制缩放比例，能控制容易分类和难分类样本的权重. 以类别 1 为例，预测概率 $f(x)$ 越大，说明该样本就越容易分类. 此时 $(1-f(x))^\gamma$ 和分类概率 $f(x)$ 成反比. 故聚焦损失相比加权交叉熵能帮助模型集中训练更加困难的样本. 超参数 α、γ 均通过贝叶斯优化[29]进行选择. 通过上述分析，聚焦损失通过 α 调整多数类和少数类样本的权重，通过 γ 控制难易分类的样本权重来解决类别不均衡问题[20]. 同时聚焦损失在 GBDT 集成学习框架下不仅能消除不均衡数据带来的影响，而且能提高模型的泛化能力和鲁棒性. 故该方法的基本步骤如下：

步骤 1　从训练集 Δ 抽取所有样本 $\{(x_i,y_i)\}_{i=1}^n$ 以及从分布 $P_Z(z)$ 随机抽取样本容量为 n 的一个样本 $\{z_i\}_{i=1}^n$；

步骤 2　将步骤 1 的数据输入到生成器 G_θ 中进而输出 n 个生成数据 $\{G_\theta(z_i\mid y_i)\}_{i=1}^n$；

步骤 3　判别器 D_α 根据步骤 2 输入的生成数据和步骤 1 中输入的数据，判别器通过随机梯度上升更新参数 α，生成器随机梯度下降更新参数 θ；

步骤 4　重复步骤 1～3 直到目标函数 $\hat{V}(D_\alpha,G_\theta)$ 的值趋于稳定，此时求得的参数为 $\hat{\theta},\hat{\alpha}$，训练 CGAN 得到的生成器为 $G_{\hat{\theta}}$；

步骤 5　从生成器 $G_{\hat{\theta}}$ 中生成 m 个少数类样本，补充到原有训练集 Δ 中，新的训练集为 Δ'. $k=1,\cdots,K$，表示迭代次数，现观察第 k 轮迭代用该轮的第 i 个样本的聚焦损失的负梯度[30]：

$$r_{ki}=-\left[\frac{\partial L(y,f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{k-1}(x)},$$

并对 r_{ki} 拟合一个回归树，通过求回归树的各个叶子节点的最优拟合值，然后更新第 k 轮的基分类器，具体过程，参考 GBDT 算法的实现；

步骤 6　对步骤 5 经过 K 次迭代更新，最后获得在训练集 Δ' 上的分类结果. 其中通过交叉验证来选择合适的 K 值.

注：①训练集在训练模型前进行归一化处理，使得每个特征值在 $[0,1]$ 之间，有助于 CGAN 模型稳定训练；②对于样本量特别大的情形，对于步骤 1 可以随机抽取容量为 t 的样本来计算梯度，提高训练速度并且通过交叉验证来选取 t 的大小；③生成器 $G_{\hat{\theta}}$ 所生成的少数类样本容量 m 是一个超参数，通过交叉验证来选取合适的 m，使得实际分类效果较好；④步骤 1 中对 y_i 进行 one-hot 编码后，与相应的 x_i 合并成一个新的向量作为隐藏层的输入，隐藏层之间通过神经元的线性组合，在非线性激活函数的作用下，对下一个隐藏层的输入即为带有标签信息的数据. 响应变量 y_i 对 z_i 合并成一个新向量操作

与 x_i 一致.

本文证明了由 CGAN 产生的经验条件分布逼近总体条件分布. 因此,CGAN 方法可以根据训练数据生成逼近真实数据分布的样本,但生成的样本不一定能完全解决类别不均衡问题. 基于这点考虑,本文在 CGAN 的基础上结合聚焦损失以及 GBDT 集成学习算法来解决不同类别不均衡的问题. 同时也证明了该方法的经验风险收敛到期望风险以及方法的估计量收敛到使得期望风险最小化的函数. 聚焦损失的 CGAN 集成分类方法将数据和算法层面相结合共同作用到不均衡数据,提高了分类效果.

本文研究方法的伪代码,描述如算法 1.1 所示.

算法 1.1　聚焦损失的 CGAN 分类方法

输入:训练数据集 Δ

输出:经过算法处理训练集的分类结果

Step 1:对训练数据进行归一化处理即 $x^* = \frac{x-\min(x)}{\max(x)-\min(x)}$,有助于模型稳定训练,处理后的数据集仍记为 Δ;

Step 2:基于训练数据集 Δ 训练 CGAN 模型:

for 训练迭代次数 do

for k 步 do

从先验均匀分布[25] U[0,1]中随机抽取 t 个随机噪声 $\{z_1, z_2 \cdots, z_t\}$;

从 Δ 随机抽取 t 个样本 $\{(x_i, y_i)\}_{i=1}^t$;

根据如下目标函数更新判别器:

$$\nabla_\alpha \frac{1}{t}\sum_{i=1}^{t}[\log D_\alpha(x_i|y_i)+\log(1-D_\alpha(G_\theta(z_i|y_i)))];$$

end for

从先验均匀分布 U[0,1]中随机抽取 t 个随机噪声 $\{z_1, z_2 \cdots, z_t\}$;

根据如下目标函数更新生成器:

$$\nabla_\theta \frac{1}{t}\sum_{i=1}^{t}\log(1-D_\alpha(G_\theta(z_i|y_i)));$$

end for

Step 3:通过模型 CGAN 得到样本容量为 m 的少数类样本补充到 Δ 中,其新的数据集为 Δ′;

Step 4:将算法的损失函数取为聚焦损失后,把数据集 Δ′带入聚焦损失的 GBDT 模型训练;

Step 5:输出训练集分类结果.

注:①在 CGAN 训练过程中,每次迭代过程中判别器会迭代次数为 k,一般来说 k 都取 1[8];②其中 ∇_α,∇_θ 分别表示判别器和生成器对未知参数求梯度.

在伪代码中,CGAN 模型通过 Adam[31] 优化算法来训练模型. 本文使用了贝叶斯优化在 Python 中的实现 Hyperopt[29] 模块进行调参. 同时,相较于 k 折交叉验证,贝叶斯优化考虑了之前的参数信息,能更好地调整当前的参数. 因此,本文对于聚焦损失超参数 α、γ,以及后文中涉及的算法参数的选取均通过贝叶斯优化进行调参. CGAN 模型中隐藏层的节点个数一般选取为输入变量维度的 2 ～ 10 倍[8],通过人工调节隐藏的个数在 3 层时表现较优.

1.3　聚焦损失的 CGAN 方法性质

设 P^* 是样本对应总体 (X,Y) 的联合分布,P_Z 为噪音变量的分布,$\mathcal{G}$ 和 $\mathcal{D}$ 如上文所示. $\mathcal{P}_{X|Y} = \{P_{\theta,X|Y}: G_\theta(Z \mid y) \sim P_{\theta,X|Y=y}, \theta \in \Theta\}$ 是包含样本对应总体的条件分布 $(P^*_{X|Y})$ 的分布族,π_{P^*} 为 Y 对应于总体 P^* 的边缘分布,$\pi_{P^*}(1)$ 表示少数类样本的概率,即 $\pi_{P^*}(1)=P^*(Y=1)$. 参考 Biau 等[32] 所证明的定理 4.1,假设对任意 $\theta \in \Theta$,$\varepsilon>0$,存在 $\beta \in (0, 1/2)$,$D \in \mathcal{D}$ 使得 $\beta \leqslant D \leqslant 1-\beta$,$\| D - D^*_\alpha \| \leqslant \varepsilon$,其中 $D^*_\alpha = \frac{dP^*_{X|Y}}{dP^*_{X|Y}+dP_{\alpha,X|Y}}$,其中 d 表示 Radon-Nikodym 导数.

定理 1.1　设 $Q_{n,X|Y}=P_{\hat{\theta},X|Y}$ 是由 CGAN 通过 $\{X_i,Y_i\}_{i=1}^n$ 得到的条件分布,在文献[32,定理 4.1]的条件下,

$$ED_{JS}(P^*_{X|Y=1}, Q_{n,X|Y=1}) \to 0,$$
$$ED_{TV}(P^*_{X|Y=1}, Q_{n,X|Y=1}) \to 0,$$

其中,D_{JS} 为 JS 散度:

$$D_{JS}(P, Q) = \frac{1}{2}D_{KL}(P \,\|\, \frac{P+Q}{2}) + \frac{1}{2}D_{KL}(Q \,\|\, \frac{P+Q}{2}),$$

D_{KL} 为 KL 散度:

$$D_{KL}(P \,\|\, Q) = E_{x\sim P}[\log \frac{dP(x)}{dQ(x)}],$$

$D_{TV}(P,Q) = \sup\limits_{A\in \text{可测集}} |P(A)-Q(A)|$ 为全变差距离.

证明　对任意的 $\varepsilon>0$ 足够小,取 $\hat{D} \in \mathcal{D}$ 使得 $\beta \leqslant \hat{D} \leqslant 1-\beta$,$\| \hat{D} - D^*_{\hat{\alpha}} \| \leqslant \varepsilon$. 按照文献[32,定理 3.1]的证明,存在常数 c_1 使得

$$2D_{JS}(P^*, Q_n) \leqslant c_1\varepsilon^2 + V(G_{\hat{\theta}}, \hat{D}) + \ln 4.$$

记 $Q_n = P_{\hat{\theta},X|Y}\pi_{P^*}(Y)$,于是按照文献[32,定理 4.1]的证明,有

$$ED_{JS}(P^*, Q_n) = O(\varepsilon^2 + \frac{1}{\sqrt{n}}),$$

先令 n 趋于无穷,再让 ε 趋于 0,则有

$$ED_{JS}(P^*, Q_n) \to 0.$$

因为 $\exists \mu$,使得 $\mu \gg P^*$,$\mu \gg Q_n$,所以

$$A_1 = D_{KL}(P^* \,\|\, \frac{P^*+Q_n}{2}) = \int p^*(x,y)\ln\frac{2p^*(x,y)}{p^*(x,y)+q_n(x,y)}d\mu(x,y) = \sum_{i=0}^{1}\int p^*(x\mid i)\pi_{P^*}(i)\times \ln\frac{2p^*(x\mid i)\pi_{P^*}(i)}{p^*(x\mid i)\pi_{P^*}(i)+q_n(x\mid i)\pi_{Q_n}(i)}d\mu(x,y) \geqslant \sum_{k=0}^{1}\int p^*(x\mid i)\pi_{P^*}(i)\ln\frac{2p^*(x\mid i)}{p^*(x\mid i)+q_n(x\mid i)}d\mu(x,y) + \sum_{k=0}^{1}\int p^*(x\mid i)\pi_{P^*}(i)\ln\frac{\pi_{P^*}(i)}{\pi_{P^*}(i)\vee\pi_{Q(i)}}d\mu(x,y) = \sum_{i=0}^{1}\pi_{P^*}(i)D_{KL}(P^*_{X|Y=i} \,\|\, \frac{P^*_{X|Y=i}+Q_{n,X|Y=i}}{2}) + (D_{KL}(\pi_{P^*} \,\|\, \pi_{Q_n}) \wedge 0),$$

同理

$$A_2 = D_{KL}(Q_n \,\|\, \frac{P^*+Q_n}{2}) \geqslant$$

$$\sum_{i=0}^{1}\pi_{Q_n}(i)D_{KL}(Q_{n,X|Y=i}\,\|\,\frac{P^*_{X|Y=i}+Q_{n,X|Y=i}}{2})+$$
$$(D_{KL}(\pi_{Q_n}\,\|\,\pi_{P^*})\wedge 0),$$

由 JS 散度的表达式可知：

$$2D_{JS}(P^*,Q_n)=A_1+A_2,$$

注意到 $\pi_{P^*}=\pi_{Q_n}$，于是

$$2D_{JS}(P^*,Q_n)\geqslant$$
$$\sum_{i=0}^{1}2\pi_{P^*}(i)D_{JS}(P^*_{X|Y=i},Q_{n,X|Y=i})\geqslant 0.$$

因为

$$ED_{JS}(P^*,Q_n)\to 0,$$

所以

$$ED_{JS}(P^*_{X|Y=1},Q_{n,X|Y=1})\to 0.$$

由 Pinsker 不等式[33]可知

$$\frac{1}{4}ED^2_{TV}(P^*_{X|Y=1}\,\|\,Q_{n,X|Y=1})=$$
$$ED^2_{TV}(P^*_{X|Y=1}\,\|\,\frac{P^*_{X|Y=1}+Q_{n,X|Y=1}}{2})\to 0,$$

所以

$$E[D_{TV}(P^*_{X|Y=1},Q_{n,X|Y=1})]\to 0.$$

定理 1.1 表明 CGAN 产生的条件分布逼近样本总体的条件分布.

得到 CGAN 生成的条件分布 $Q_{n,X|Y=1}$ 后，目标函数 $R_{n,m}(f)=\frac{1}{m+n}\sum_{i=1}^{n}L(f(X_i),Y_i)+\frac{1}{m+n}\sum_{j=1}^{m}L(f(\widetilde{X}_{n,j}),1)$，$\mathscr{F}$ 为分类模型训练分类器的取值范围，在 $\mathscr{F}$ 中求使得目标函数最小的 f. 其中实际观测样本 $\{X_i,Y_i\}_{i=1}^n$ 独立同分布于 P^*，CGAN 生成的样本 $\{\widetilde{X}_{n,j}\}_{j=1}^m$ 来自于分布 $Q_{n,X|Y=1}$，L 聚焦损失. 假设 $\varepsilon\in(0,\frac{1}{2})$，使得对任意 $f\in\mathscr{F}$，有 f 连续，且 f 满足 $\varepsilon<f<1-\varepsilon$.

定理 1.2 对 $\forall\lambda\in[0,1)$，当 $m_n\to\infty$，使得 $\frac{n}{n+m_n}\to\lambda$，对 $\forall f\in\mathscr{F}$，有

$$H_1=E_{P^*}(L(f(X),Y)\mid Y=1),$$
$$H_0=E_{P^*}(L(f(X),Y)\mid Y=0),$$
$$R_{n,m}(f)\xrightarrow{P}\lambda\pi_{P^*}(1)+1-\lambda)H_1+\lambda\pi_{P^*}(0)H_0.$$

证明 因为，对 $\forall f\in\mathscr{F}$，有 $\varepsilon<f<1-\varepsilon$，由上述等式(2)可知，$L(f(x),y)$ 一致有界，不妨令其为 S，即 $\sup_{f\in\mathscr{F}}L(f(x),y)\leqslant S$.

由弱大数律可知

$$\frac{1}{n}\sum_{i=1}^{n}L(f(X_i),Y_i)\xrightarrow{P}\pi_{P^*}(1)H_1+\pi_{P^*}(0)H_0,$$

记

$$T_{m,n}=\frac{1}{m}\sum_{j=1}^{m}L(f(\widetilde{X}_{n,j}),1),$$
$$T_n=E_{Q_{n,X|Y=1}}(L(f(X),Y)\mid\{X_i,Y_i\}_{i=1}^n),$$

则

$$P(|T_{m,n}-T_n|>\varepsilon)\leqslant$$
$$\frac{1}{\varepsilon^2}E\mathrm{Var}_{Q_{n,X|Y=1}}(T_{m,n}\mid\{X_i,Y_i\}_{i=1}^n)\leqslant\frac{2S^2}{\varepsilon^2 m}.$$

取 $\alpha>0$，$m_n=\begin{cases}\frac{1-\lambda}{\lambda}n, & \text{if }\lambda>0;\\ n^{1+\alpha}, & \text{if }\lambda=0;\end{cases}$ $\varepsilon_n=n^{-\frac{1}{3}}$，则 $\varepsilon_n\downarrow 0$，

$$P(|T_{m_n,n}-T_n|>\varepsilon_n)\leqslant\frac{2S^2}{\varepsilon_n^2 m_n}\to 0,$$

所以 $T_{m_n,n}-T_n\xrightarrow{P}0$.

由定理 1.1 可知

$$ED_{TV}(P^*_{X|Y=1},Q_{n,X|Y=1})=$$
$$E\sup_{\|g\|_\infty\leqslant 1}|E_{P^*_{X|Y=1}}g-E_{Q_{n,X|Y=1}}g|\to 0,$$

$$P(|T_n-H_1|>\varepsilon)\leqslant\frac{1}{\varepsilon}E|T_n-H_1|\leqslant$$
$$\frac{1}{\varepsilon}S\cdot ED_{TV}(P^*_{X|Y=1},Q_{n,X|Y=1})\to 0,$$

所以

$$T_n-H_1\xrightarrow{P}0.$$

所以

$$\frac{1}{m}\sum_{j=1}^{m}L(f(\widetilde{X}_{n,j}),1)\xrightarrow{P}H_1.$$

定理 1.2 表明聚集损失的 CGAN 方法的经验风险收敛到期望风险.

令 $J(S)=\sup_Q\int_0^S\sqrt{1+\log N(\mathscr{H},L_2(Q),\varepsilon)}\,\mathrm{d}\varepsilon$.

其中，$\mathscr{H}=\{L(f(\cdot),\cdot):f\in\mathscr{F}\}$，$\mathscr{F}_0\xlongequal{\triangle}\operatorname{argmin}_{f\in\mathscr{F}}R(f)$，

$R(f)\xlongequal{\triangle}(\lambda\pi_{P^*}(1)+1-\lambda)E_{P^*}(L(f(X),Y)\mid Y=1)+\lambda\pi_{P^*}(0)E_{P^*}(L(f(X),Y)\mid Y=0)$，

$$C=\{f\mid f:\mathscr{X}\to(\varepsilon,1-\varepsilon),f\text{ 连续}\},$$
$$f_{n,m}\in\operatorname*{argmin}_{f\in\mathscr{F}}R_{n,m}(f).$$

记 $e^P_{n,2}(f,g)=\|f-g\|_{L^2(P_n)}$.

定理 1.3 设 $J(S)<\infty$，对 $\forall\lambda\in[0,1]$，$\exists m_n\to\infty$，使得 $\frac{n}{n+m_n}\to\lambda$，对 $\forall(C,\|\ \|_\infty)$ 中包含 $\mathscr{F}_0$ 的开集 U，有

$$P(f_{n,m}\in U)\to 1.$$

证明 由 M-估计的相合性可知[34]，只需证明

$$\|R_{n,m}-R\|_{\mathscr{F}}\xrightarrow{P}0.$$

$$\|R_{n,m}-R\|_{\mathscr{F}}\leqslant\|\frac{1}{m+n}\sum_{i=1}^{n}L(f(X_i,Y_i))-$$

$$\lambda E_{P^*}(L(f(X),Y))\|_{\mathscr{F}}+$$
$$\left\|\frac{1}{m+n}\sum_{j=1}^{m}L(f(\widetilde{X}_{n,j}),1)-\right.$$
$$\left.\frac{m}{m+n}E_{Q_{n,X|Y=1}}(L(f(X),1)\mid\{X_i,Y_i\}_{i=1}^{n})\right\|_{\mathscr{F}}+$$
$$\left\|\frac{m}{m+n}E_{Q_{n,X|Y=1}}(L(f(X),1)\mid\{X_i,Y_i\}_{i=1}^{n})-\right.$$
$$\left.(1-\lambda)E_{P^*}(L(f(X),Y)\mid Y=1)\right\|_{\mathscr{F}}\xlongequal{\triangle}$$
$$(i)+(ii)+(iii).$$

由假设可知

$$E_{P^*}\left(1\wedge\frac{1}{\sqrt{n}}\int_0^{2S}\sqrt{\log N(\mathscr{H},e_{n,2}^{P^*},\varepsilon)}\,\mathrm{d}\varepsilon\right)\to 0,$$

按照文献[35，定理 3.7.14]有 $\mathscr{H}$ 是 P^*-Glivenko-Cantelli class，所以 $(i)\xrightarrow{\mathrm{P}}0$. 记

$$T_{m,n}=\frac{1}{m}\sum_{j=1}^{m}L(f(\widetilde{X}_{n,j}),1),$$
$$T_n=E_{Q_{n,X|Y=1}}(L(f(X),Y)\mid\{X_i,Y_i\}_{i=1}^{n}),$$

由文献[34，定理 2.14.1]有

$$P(\|T_{m,n}-T_n\|_{\mathscr{F}}>\varepsilon)\leqslant$$
$$\frac{1}{\varepsilon}E\|T_{m,n}-T_n\|_{\mathscr{F}}\lesssim\frac{1}{\varepsilon\sqrt{m}}J(S).$$

对 m_n,ε_n 的取值和定理 1.2 中相同，有

$$P(\|T_{m_n,n}-T_n\|>\varepsilon_n)\lesssim\frac{1}{\varepsilon_n\sqrt{m_n}}J(S)\to 0,$$

所以 $(ii)\xrightarrow{\mathrm{P}}0$. 因为

$$\sup_{g\in\mathscr{H}}\|g\|_{\infty}<S,$$
$$|E_{Q_{n,X|Y=1}}(L(f(X),1)\mid\{X_i,Y_i\}_{i=1}^{n})-$$
$$E_{P^*}(L(f(X),Y)\mid Y=1)|\leqslant$$
$$S\cdot D_{TV}(Q_{n,X|Y=1},P_{X|Y=1}^{*}),\ \text{a.s.}.$$

由定理 1.1 可知 $(iii)\xrightarrow{\mathrm{P}}0$. 故

$$P(f_{n,m}\in U)\to 1.$$

定理 1.3 表明本文提出的方法的估计量会收敛到期望风险最小化对应的函数.

2　数值实验

2.1　实验数据与设计

本次研究采用的数据来自某银行信用卡中心提供的用户行为数据，模型所用的变量包括客户的基本信息以及征信信息. 为了证明模型具有普遍适用性，选取公共数据集 KEEL 的 4 个不均衡数据集进行实验. 其中定义类别为 0 和 1，其中 1 表示为少数类，0 表示多数类. 数据集的基本信息如表 1 所示.

设置 CGAN 迭代 1200 次，步长为 0.00001，采用 Adam 优化算法训练模型. 对于样本量较大的数据集，取 80%的样本量作为训练集，20%作为测试集. 对样本量较小的数据集采用 k 折交叉验证. 模型使用贝叶斯优化进行调参. 通过聚焦损失的 CGAN 方法在数据集上的表现来验证其有效性.

表 1　实验数据集的基本信息

Fig. 1　Basic information of the experimental dataset

数据集	样本数	样本分布	不均衡比率	特征数
信用卡数据	67773	1722,66051	38.4	45
ecoil3	336	35,301	8.6	7
yeast1	1484	429,1055	2.5	8
yeast4	1433	51,1382	27.1	8
cleveland-0-vs-4	173	13,160	12.3	13

由于算法层面的基分类器是二分类的决策树，因此为了验证 CGAN 作为抽样方法的作用，选用决策树模型进行比较. CGAN 作为抽样方法将和各种重抽样方法进行对比，以 AUC 值为评判标准，结果显示出 CGAN 抽样方法的优势. 之后固定 CGAN 作为抽样方法，通过实验来说明算法层面上聚焦损失的 GBDT 具有一定的优势. 最后将 LR(logistic regression)，RF(random forest)，KNN，XGBoost 与各种抽样方法相结合与聚焦损失的 CGAN 进行对比，以 AUC 值为评判标准，显示了该算法的有效性.

2.2　信用卡数据集实证分析

首先，通过 CGAN 对训练集的少数类样本进行数据生成. 原理是判别器 D 和生成器 G 两个网络交替训练，使得 G 产生的数据能够尽量迷惑 D，而 D 可以分辨真实样本和假样本. 所以 G 产生的数据分布会越来越接近真实的数据分布，D 的辨别能力也会越来越强，最终到达一个平衡. 从图 1 中可以看出，两个网络交替迭代 1500 次，通过观察两个网络的损失值可以发现，模型训练较好.

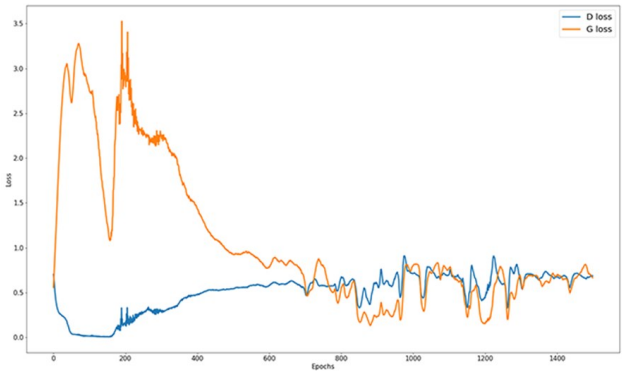

图 1　生成器和判别器的损失值

Fig. 1　Generator and discriminator loss values

以某银行信用卡用户的年龄 PAI_AGE 和月收入 PAI_MONTH_INCOME 这两个属性为例，绘制真实样本的分布图和生成数据的分布图之间的差异. 图 2 表示了迭代 500 次、1000 次、1500 次时真实样本和生成样本分布的差异. 从图 2 可以看出，随着迭代次数的增加，CGAN 模型生成的样本分布和观测样本的分布之间的差异越来越小.

由以上图 1 和 2 可以看出，CGAN 生成的数据质量较好. 将 CGAN 和重抽样方法 ROS、SMOTE、borderline SMOTE、ADASYN、RUS、SMOTEENN、GAN 分别与决策树进行结合，并对比其分类效果. 对比结果如表 2 所示.

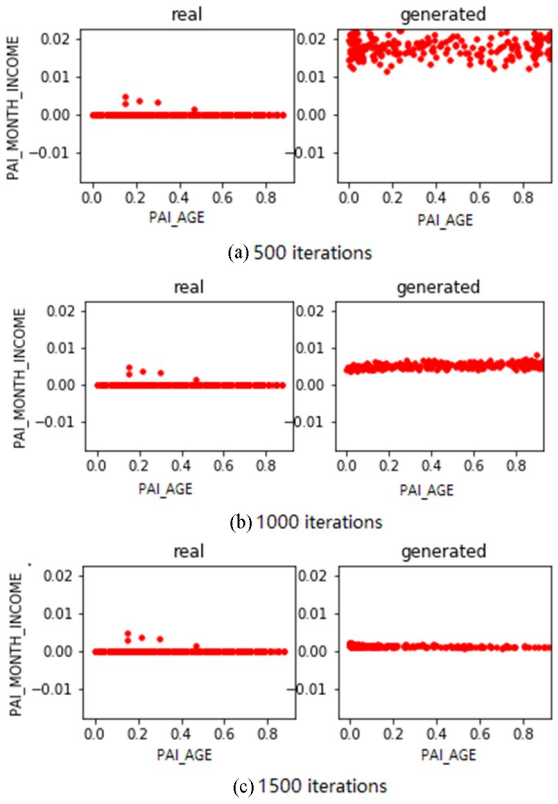

图 2 迭代 500 次,1000 次,1500 次真实数据和生成数据的经验分布

Fig. 2 500, 1000, 1500 iterations of real data and empirical distribution of generated data

表 2 各种抽样方法的 AUC 值

Tab. 2 AUC values of various sampling methods

算法	AUC	training AUC
ROS-DT	0.6928	0.6902
SOMTE-DT	0.6508	0.7188
broderline SOMTE-DT	0.6379	0.7554
ADASYN-DT	0.6508	0.7132
SMOTEENN-DT	0.6867	0.7469
RUS-DT	0.6641	0.6640
GAN-DT	0.7056	0.8219
CGAN-DT	**0.7393**	**0.8341**

从表 2 可以看出,CGAN 作为数据重抽样方法,其 AUC 的值最高,并且通过对比 AUC 与 training AUC 的结果可以知道,各种抽样方法,经过调参得到比较合适的结果. 与欠抽样方法 RUS 相比,CGAN 的 AUC 值提高 11.32%,相对于过抽样方法中表现最好的 ROS,其 AUC 值提高 6.711%,相对于其他抽样的方法 SMOTEENN,其 AUC 值提高 7.65%,相对 GAN 方法 AUC 值提高 4.77%. 从 AUC 维度来看,新的抽样方法 CGAN 整体来说表现较为优异.

表 3 以 CGAN 作为抽样方法,来说明在算法层面上聚焦损失的 GBDT 方法具有一定的优势. 从表 3 可以看出聚焦损失的 CGAN 表现最优. 与算法中仅使用 GBDT 方法相比 AUC 值提高了 2.39%;该方法与 LR,RF,XGBOOST,KNN 其 AUC 值分别提升了 5.04%,3.76%,4.24%,9.10%.

表 3 各种分类方法的 AUC 值

Tab. 3 AUC values of classification methods

算法	AUC	training AUC
focal loss-GBDT	**0.7914**	0.8764
GBDT	0.7729	0.8241
LR	0.7534	0.8945
RF	0.7627	**0.8984**
XGBOOST	0.7592	0.8851
KNN	0.7254	0.8494

表 4 为 LR,RF,KNN,XGBoost 与各种抽样方法相结合与聚焦损失的 CGAN 算法在信用卡数据集上所得出的 AUC 值.

表 4 各种分类算法的 AUC 值

Tab. 4 AUC values of classification algorithms

算法		AUC	training AUC
聚焦损失的 GBDT	GAN	0.7818	0.8518
	CGAN	**0.7914**	**0.8764**
LR	GAN	0.7663	**0.8661**
	ROS	**0.7685**	0.7728
	SMOTE	0.7651	0.7843
	borderline SMOTE	0.7637	0.8571
	ADASYN	0.7638	0.7743
	SMOTEENN	0.7655	0.8118
	RUS	0.7606	0.7569
RF	GAN	0.7598	0.8787
	ROS	0.7547	0.8672
	SMOTE	0.7295	0.9602
	borderline SMOTE	0.7339	**0.9705**
	ADASYN	0.7199	0.9514
	SMOTEENN	0.7390	0.9601
	RUS	**0.7677**	0.7615
XGBoost	GAN	**0.7574**	0.8884
	ROS	0.7491	**0.9999**
	SMOTE	0.7053	0.9956
	borderline SMOTE	0.7186	0.9947
	ADASYN	0.7009	0.9955
	SMOTEENN	0.7182	0.9964
	RUS	0.7447	0.7690
KNN	GAN	0.6455	0.8212
	ROS	0.5607	0.9873
	SMOTE	0.5999	0.9837
	borderline SMOTE	0.5815	0.9856
	ADASYN	0.5768	0.9653
	SMOTEENN	0.6059	**0.9882**
	RUS	**0.7161**	0.7133

从表 4 可以看出聚焦损失的 CGAN 在信用卡

数据分析中综合表现最优. 与 LR 中表现最好的抽样方法 ROS 相比,该方法的 AUC 值提高 3.00%;与 RF 中表现最好的抽样方法 RUS 相比,其 AUC 值提高 3.09%;与 XGBoost 中表现最好的抽样方法 GAN 相比,其 AUC 值提高 4.49%;与 KNN 中表现最好的抽样方法 RUS 相比,其 AUC 值提高了 10.52%.

2.3　其他数据集实证分析

通过选取 KEEL 的 4 个数据集进行数值实验,来进一步说明聚焦损失的 CGAN 适用性比较强. 目前为止使用的重抽样方法几乎都是针对某一类的样本,抽样方法存在一定的局限性. 因此,采用方法选取上文提到的 SMOTEENN 方法来进行对比,聚焦损失的 CGAN、DT、LR、RF、XGBoost、KNN 在这些数据集上所得的 AUC,training AUC 值如表 5 所示. 其中聚焦损失的 CGAN 简写为 FLCGAN.

表 5　各个数据集的 AUC 值

Tab. 5　AUC values of data sets

算法		ecoil3	yeast1	yeast4	cleveland-0-vs-4
FLCGAN	AUC	**0.9532**	**0.8108**	**0.8899**	**0.9896**
	training AUC	0.9614	0.8637	0.7859	0.9628
DT	AUC	0.9344	0.6959	0.7923	0.6667
	training AUC	0.9841	0.9095	0.9734	0.9465
LR	AUC	0.9415	0.7795	0.8745	0.9688
	training AUC	0.9984	0.9117	0.9494	**1.0000**
RF	AUC	0.9403	0.7806	0.8472	0.8958
	training AUC	0.9983	0.9809	0.9736	1.0000
XGBoost	AUC	0.9344	0.7711	0.8507	0.9375
	training AUC	0.9989	**0.9827**	**0.9984**	0.9993
KNN	AUC	0.8864	0.7873	0.8734	0.9375
	training AUC	**0.9996**	0.9782	0.9975	0.9957

从表 5 可以看出,这四个数据集中聚焦损失的 CGAN 表现性能最好,与其他算法对比,该四个数据集 AUC 值最低都分别提高了 1.24%,2.98%,1.76%,5.56%. 从表 5 的结果可以看出,聚焦损失的 CGAN 具有一定的普遍适用性,并且表现较好.

3　结论

本文针对类别不均衡数据集,从数据和算法层面研究了聚焦损失的 CGAN 方法,该方法主要将前面两种层面进行整合,以优化算法的分类效果. 该方法通过 CGAN 生成数据来降低数据的不均衡率. 从算法层面,通过权值均衡的聚焦损失可以使模型在训练过程中更加关注少数类样本和难分类样本,从而提高分类算法的分类性能,并且通过 GBDT 集成学习提高模型整体的泛化能力.

将 CGAN 模型和聚焦损失应用在数据和算法层面来处理类被不均衡数据,拓宽了解决问题的方法,同时该方法仍存在较大的提升空间. 因为 CGAN 中的判别器 D 用来评估真实样本分布和生成数据分布之间的差异或距离,所以在训练 D 时可以用密度比估计,从密度比估计的角度来进一步学习深度生成模型以便于提高模型训练的稳健性. 因此下一步的研究重点是:将密度比估计方法带入到 GANs 相关方法中,从密度比估计的维度进一步拓展生存对抗网络的宽度;解决 CGAN 模型超参数的优化方法;利用 CGAN 和聚焦损失来解决不均衡的多分类问题.

参考文献(References)

[1] AKBANI R, KWEK S, JAPKOWICZ N. Applying support vector machine to imbalanced datasets[C]// Machine Learning: ECML 2004. Berlin: Springer, 2004: 39-50.

[2] MAZUROWSKI M A, HABAS P A, ZURADA J M, et al. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance[J]. Neural Networks, 2008, 21:427-436.

[3] TAVALLAEE M, STAKHANVA N, GHORBANI A A. Toward credible evaluation of anomaly-based intrusion-detection methods[J]. IEEE Transactions on Systems, Man, and Cybernetics Part C, 2010, 40(5):516-524.

[4] BERMEJO P, GSMEZ J A, PUERTA J M. Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets[J]. Expert Systems with Applications, 2011, 38(3): 2072-2080.

[5] WEI W, LI J, CAO L, et al. Effective detection of sophisticated online banking fraud on extremely imbalanced data[J]. World Wide Web, 2013, 16: 449-475.

[6] KERDPRASOP K, KERDPRASOP N. A data mining approach to automate fault detection model development in the semiconductor manufacturing process[J]. International Journal of Mechanics, 2011, 5(4): 336-344.

[7] CHAWLA N V, JAPKOWICZ N, KOTCZ A. Editorial: Special issue on learning from imbalanced data sets[J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 1-6.

[8] DOUZAS G, BACAO F. Effective data generation for imbalanced learning using conditional generative adversarial networks[J]. Expert Systems with Applications, 2018, 91: 464-471.

[9] CHAWLA N V. Data mining for imbalanced datasets: An overview[C]// Data Mining and Knowledge Discovery Handbook. Berlin: Springer, 2009: 875-886.

[10] 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(1): 321-357.

[11] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning[J]. ICIC 2005: Advances in Intelligent Computing, 2005, 17(12): 878-887.

[12] HE H, BAI Y, GARCIA E A, et al. ADASYN:

Adaptive synthetic sampling approach for imbalanced learning[C]// 2008 IEEE International Joint Conference on Neural Networks. IEEE, 2008: 1322-1328.
[13] BATISTA G E, PRATI R C, MONARD M C. A study of the behavior of several methods for balancing machine learning training data[J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 20-29.
[14] FAN W, ZHANG J, STOTFO S J, et al. AdaCost: Misclassification cost-sensitive boosting [C]// Proceedings of the 16th International Conference on Learning, Slovenia: Morgan Kaufmann, 1999: 97-105.
[15] WOZNIAK M. Classifiers: Methods of data, knowledge, and classifier combination[M]. Berlin: Springer, 2013.
[16] MARIANI G, SCHEIDEGGER F, ISTRATE R, et al. BAGAN: Data augmentation with balancing GAN [DB/OL]. [2020-05-01]. https://arxiv.org/abs/1803.09655.
[17] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [DB/OL]. [2020-05-01]. https://arxiv.org/abs/1511.06434v1.
[18] GAO Y, JIAO Y, WANG Y, et al. Deep generative learning via variational gradient flow[DB/OL]. [2020-05-01]. https://arxiv.org/abs/1901.08469.
[19] ZHANG Y. Deep generative model for multi-class imbalanced learning[DB]// Open Access Master's Theses, 2018: Paper 1277.
[20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 2980-2988.
[21] FAWCETT T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.
[22] 赵海霞,石洪波,武建,等. 基于条件生成对抗网络的不平衡学习研究[J/OL]. 控制与决策, 2019: https://doi.org/10.13195/j.kzyjc.2019.0522.
[23] 莫赞,盖彦蓉,樊冠龙. 基于 GAN-AdaBoost-DT 不平衡分类算法的信用卡欺诈分类[J]. 计算机应用, 2019, 39(2):618-622.
[24] 李诒靖,郭海湘,李亚楠,等. 一种基于 Boosting 的集成学习算法在不均衡数据中的分类[J]. 系统工程理论与实践, 2016, 36(1):189-199.
[25] GOODFELLOW I, POUGET A J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems Conference, 2014, 27: 2672-2680.
[26] SALANT S W, SWITZER S, REYNOLDS R J. Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium[J]. The Quarterly Journal of Economics, 1983, 98(2):185-199.
[27] MIRZA M, OSINDERO S. Conditional generative adversarial nets [C]// Proceedings of the Neural Information Processing Systems Deep Learning Workshop, 2014.
[28] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
[29] BERGSTRA J, YAMINS D, DAVID D C. Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms[C]// Proceedings of the 12th Python in Science Conference, Austin, TX, 2012.
[30] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
[31] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]// The 3rd International Conference on Learning Representations, San Diego, CA, 2015.
[32] BIAU G, CADRE B, SANGNIER M, et al. Some theoretical properties of GANs[DB/OL]. [2020-05-01]. https://arxiv.org/abs/1803.07819.
[33] TSYBAKOV A B. Introduction to Nonparametric Estimation[M]. Berlin: Springer, 2008.
[34] VAN DER VAART A W, WELLNER J. Weak Convergence and Empirical Processes[M]. Berlin: Springer, 2000.
[35] GIN W E, NICKL R. Mathematical Foundations of Infinite Dimensional Statistical Models [M]. Cambridge: Cambridge University Press, 2015.
[36] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.