基于数据增强与 XGBoost 模型的信用卡欺诈检测算法

邓磊
合肥工业大学 数学科学学院，安徽 合肥 230009
DOI:10.61369/ASDS.2026040007

摘　　要： 信用卡欺诈行为对经济、金融等诸多领域产生严重不良影响，如何有效预防、识别信用卡欺诈行为具有重要理论和实际价值。另一方面，来自银行匿名的真实数据集，高度不平衡，正常数据远远大于欺诈数据。本文首先基于 CGAN 和 Isolation Forest 对数据加强，构建贝叶斯优化器的 XGBoost 模型，用于检测信用卡欺诈，称为 CGAN–IF–Bayes–XGB 算法。具体来说使用 CGAN 生成新的、真实的数据来补充原始数据集，Isolation Forest 则用于计算每条样本的“异常性”评分，作为新的特征注入，以此来增加样本中的有效信息量，为 XGBoost 模型更好地识别诈骗交易做好铺垫。在 CGAN–IF–Bayes–XGB 算法中使用了贝叶斯优化器来对 XGBoost 中的超参数进行优化，确保 XGBoost 识别效果达到最佳。其次，本文把 CGAN–IF–Bayes–XGB 算法与其他常用的机器学习算法（例如 KNN 和 Light GBM）进行了比较，比较结果表明：CGAN–IF–Bayes–XGB 算法具有更高的准确率、真阳性率、真阴性率和马太相关系数，这使其成为检测信用卡欺诈的有效方法。

关 键 词： CGAN 模型；XGBoost 模型；Isolation Forest 模型；贝叶斯优化器

Credit Card Fraud Detection Algorithm Based on Data Augmentation and XGBoost

Deng Lei
School of Mathematical Sciences, Hefei University of Technology, Hefei, Anhui 230009

Abstract: Credit card fraud detection is challenged by severe class imbalance in real–world banking datasets. This study proposes a hybrid framework, termed CGAN–IF–Bayes–XGB, which integrates Conditional Generative Adversarial Networks (CGAN), Isolation Forest (IF), Bayesian Optimization, and XGBoost for fraud detection. CGAN is used to generate synthetic fraudulent samples to alleviate data imbalance, while IF provides anomaly scores as auxiliary features to enhance discriminative representation. Bayesian Optimization is applied to automatically tune XGBoost hyperparameters for optimal classification performance. Experimental results on a real–world anonymized banking dataset show that the proposed model outperforms baseline methods, including KNN and LightGBM, in terms of accuracy, TPR, TNR, and MCC, demonstrating its effectiveness and robustness in credit card fraud detection.

Keywords: CGAN model; XGBoost model; Isolation Forest model; Bayesian optimization

引言

随着电子支付和在线交易的快速普及，信用卡支付成为全球范围内主要的非现金支付方式。然而，信用卡的广泛应用也使其欺诈问题日益突出，这也成为金融机构面临的重大风险来源[1]。相关研究和行业报告表明，全球每年因信用卡欺诈造成的经济损失高达数百亿美元，且呈现出逐年上升的趋势[2]。在中国，信用卡日益普及、金融电子化加速发展，信用卡累计发行量已突破数亿张，全国范围内交易金额庞大，市场规模巨大，这也给金融风险管理带来严峻挑战[3]。

在上述背景下，欺诈性的信用卡交易行为如何被准确识别出来已经成为风险控制领域的核心研究课题。相较于传统的规则匹配或人工审核，数据驱动的自动化欺诈检测能够在大规模交易数据中实时识别潜在的风险交易，这种方法已被金融机构广泛应用[4]。然而，在实际应用中，欺诈的交易行为往往有着隐蔽性强、模式多变以及与正常交易高度相似等特点，这使得欺诈检测面临较大挑战[1]。

基于上述分析，本文提出了一种基于条件生成对抗网络（CGAN）的数据增强、加入 *Isolation Tree* 异常特征评分以及使用贝叶斯优化 *XGBoost* 模型的新型信用卡欺诈检测方法 (CGAN–IF–Bayes–XGB)。该方法从数据层、特征层和模型层三个层面对传统欺诈检测流程进行改进：首先，利用 CGAN 学习真实欺诈交易的分布特征，生成高质量的欺诈样本以缓解交易数据高度不平衡的问题[5、6]；其次，使用 *Isolation Tree* 对 CGAN 生成的欺诈样本进行质量评估与筛选，将高质量的生成数据与原始训练集进行融合[5]；同时，基于 *Isolation*

作者简介：邓磊，合肥工业大学数学科学学院，硕士，研究方向为应用统计。

Tree 的异常检测对交易行为进行建模，并将其输出的异常评分作为辅助特征引入后续模型 [7-9]，以增强对欺诈交易的判别能力；最后，采用贝叶斯优化器对 *XGBoost* 分类模型中关键的超参数进行自动搜索，以提升模型整体性能和稳定性 [10、11]。

针对信用卡交易数据高度不平衡和欺诈样本特征难以表达等问题，本文的主要创新性体现在如下几个方面：一是提出了一种融合条件生成和异常检测的 CGAN-IF-Bayes-XGB 集成方法。将 CGAN 用于欺诈数据增强、*Isolation Tree* 用于异常性特征注入，并结合贝叶斯优化的 *XGBoost* 分类器，形成一个完整有效的欺诈检测方法。二是使用 CGAN 生成高质量的欺诈样本以缓解数据不平衡问题。与传统过采样方法相比，CGAN 能够学习真实欺诈样本的分布，从而生成更自然、更加多样的新欺诈样本。三是引入 *Isolation Tree* 异常评分作为附加特征以增强区分能力 [7、8]。对每条交易计算异常得分，并作为新特征融合进训练数据，使分类器能够更好地捕获潜在欺诈行为的异常性，提高识别准确率。四是采用贝叶斯优化自动搜索 *XGBoost* 的最优超参数。相比于网格搜索和随机搜索，贝叶斯优化不仅搜索效率更高，而且能更稳定地找到性能最优的超参数组合。

一、文献综述

传统的信用卡欺诈检测方法主要包括基于规则的系统、统计方法以及机器学习（Machine Learning，ML）方法。基于规则的系统依赖预先定义的规则集，主要来源于业务经验和专家知识，无法处理新型的欺诈模式，误报率较高，且规则维护依赖人工，系统灵活性不足 [4]。统计方法对交易数据的统计指标进行建模，如均值、中位数、标准差及聚类分析等，用于识别异常交易。这类方法通常基于分布假设，在处理复杂、特征维度较高的信用卡交易数据时适用性有限 [4]。随着数据规模和计算能力的提升，机器学习方法被广泛应用于信用卡欺诈检测，通过从历史数据中学习非线性特征来提高检测性能。然而，在高度不平衡的数据环境下，传统机器学习模型仍易受到类别分布不均衡的影响，导致难以识别欺诈性交易 [12]。

在检测信用卡欺诈问题中主要挑战是数据分布不平衡，即与合法交易相比，欺诈交易非常少，这可能会导致检测模型有偏差。国内外许多研究人员都尝试使用机器学习和深度学习算法来检测信用卡欺诈。他们为了解决数据不平衡的问题广泛使用 SMOTE 算法来确保训练数据集中正样本和负样本的平衡表示 [13]，他们概述了各种采样技术及其不同的方法，还研究了不平衡的数据如何影响学习算法的性能，从而导致不准确、结果不正确以及 F1 评估、召回率和精确率分数降低。

近年来，机器学习方法被广泛应用于异常检测中，并在金融欺诈、网络安全和工业故障诊断等领域取得了很大成就。Nassif 等人对机器学习异常检测方法进行了系统综述，指出在高维、类别高度不平衡的数据场景下，无监督异常检测方法具有显著优势。其中，*Isolation Tree* 由于无需依赖数据分布、计算效率高、能够有效识别异常样本以及能够输出连续型异常评分而被广泛应用，其生成的异常评分还可作为辅助特征引入后续分类模型，从而提升整体检测性能 [1]。

近期，Choi et al. 将 CGAN 模型应用于信用卡数据集上生成新的交易数据 [14]。CGAN 会生成一个信用卡数据集，该数据集可以在不透露实际数据集的情况下用于训练。他们的研究结果表明，深度学习模型可以应用于信用卡交易数据生成领域。

Lee 和 Par 认为，在处理大量数据集时，深度学习模型的性能优于机器学习模型。作者认为，以往在关于解决不平衡类挑战的研究中，由于过度拟合和数据丢失而存在局限性。为了克服这些问题，他们利用 GAN 和他们提出的模型来处理不平衡的类。然后，他们使用随机森林分类评估模型的检测能力，并基于 GAN 进行数据增强。结果表明，他们提出的模型优于不考虑不平衡类的模型。

总的来说，现有研究在信用卡欺诈异常检测中取得了一定成果，但欺诈检测仍面临诸多挑战，例如交易数据中有效信息密度较低，模型在不同时间段下的稳定性和泛化能力不足、在大规模数据环境下的计算效率不足。针对上述问题，相关研究普遍认为，通过融合数据生成、异常检测与监督分类的多阶段建模策略，可在一定程度上缓解数据不平衡和特征表达受限等问题 [15、16]。

二、模型与方法

本文提出了一种融合数据生成、异常检测与监督分类的信用卡欺诈检测方法，称之为 CGAN-IF-Bayes-XGB。该方法首先使用 CGAN 生成欺诈的交易样本 [5、6]，并通过 *Isolation Tree(IF)* 筛选出高质量的欺诈数据，以缓解样本类分布不平衡的问题。然后，将筛选后的合成数据与原始训练集数据混合，利用 IF 计算交易异常评分，并将其作为新特征引入数据集，以增强特征表达能力 [5]。最后，在训练集中训练 *XGBoost* 分类模型，并使用贝叶斯优化器对模型的超参数进行调节，通过 AUC 指标选取最优超参数作为最终的欺诈检测模型 [10、17]。整体流程如图 1 所示：

（一）条件生成对抗网络（CGAN）

传统 GAN 由生成器和判别器构成，通过博弈进行联合优化：生成器 G：生成与真实数据分布近似的数据。判别器 D：区分数据是来自真实数据集还是生成器。其目标函数可表示为 [6]：

$$\underset{G}{MIN}\ \underset{D}{MAX}\ V(D,G)=E_{x\sim P_{data}}[\log D(x)]+E_{z\sim P_z}[\log(1-D(G(z)))] \quad (1)$$

其中：$x\sim P_{data}(x)$ 表示真实样本 x 服从于真实样本 $P_{data}(x)$ 分布，即

从原始信用卡交易样本集中随机采样得到的交易样本；$z \sim P_z$ 是从正态分布中采样的随机向量；$G(z)$ 是生成器生成的数据；$D(x)$ 是判别器判断 x 是来自真实样本的概率。

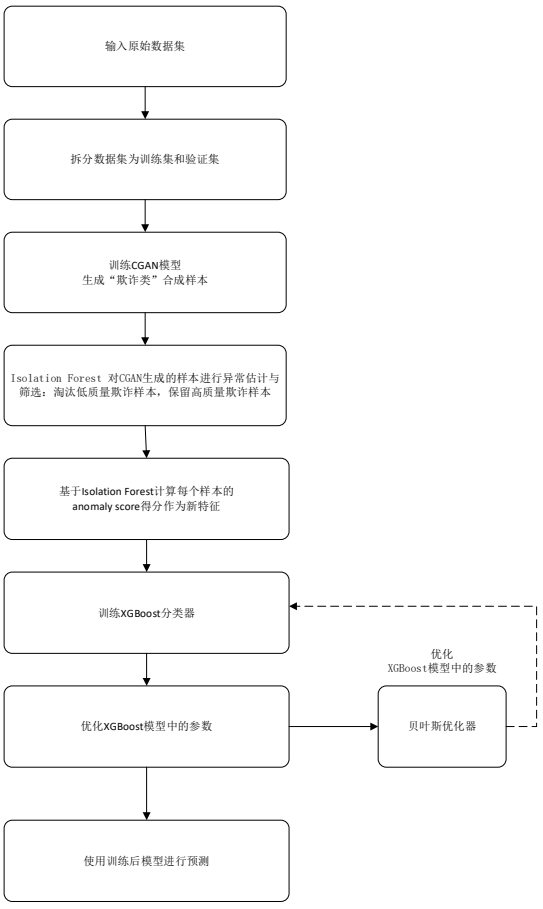

图1：基于 CGAN-Isolation Forest-XGBoost 的信用卡欺诈检测模型流程图

CGAN 是将条件变量 y（如类别、标签、特征）同时引入生成器和判别器，生成特定类型的样本 [18、19]。比如在信用卡欺诈场景中，将 y=0 代表正常交易，y=1 代表欺诈交易，作为条件变量控制生成样本的类型 [20-22]。CGAN 的目标函数如下 [23]：

$$\underset{G}{MIN}\underset{D}{MAX}V(D,G)=E_{x\sim P_{data}}\left[\log D(x|y)\right]+E_{z\sim P_z}\left[\log\left(1-D(G(z|y))\right)\right] \quad (2)$$

在生成器中，将 z 与条件 y 一起拼接，输入网络：$G(z|y)=G([z,y])$。

在判别器中，将生成/真实样本 x 与条件 y 一起输入：$G(x|y)=G([x,y])$。

在模型训练过程中，CGAN 通过交替优化生成器与判别器，使二者在对抗中逐步收敛。判别器在给定条件标签的约束下学习区分真实交易样本与生成样本，而生成器则在随机向量与条件信息的共同驱动下，不断调整参数以生成更接近真实欺诈样本分布的样本。随着训练的进行，生成器能够逐步识别欺诈交易的潜在特征，从而生成满足条件约束且分布一致性较高的欺诈样本，为后续异常筛选与分类建模提供可靠的数据基础 [19]。

在本文提出的 CGAN-IF-Bayes-XGB 框架中，CGAN 主要用于生成条件为 y=1 的欺诈交易样本，以缓解原始数据集中缺少欺诈交易样本的问题。

（二）Isolation Forest 异常检测模型

虽然 CGAN 能够生成与真实样本分布相似的样本，但其生成样本的质量仍存在不确定性，部分样本可能偏离真实分布且引入噪声 [6]。为此，本文进一步引入 *Isolation Tree(IF)* 对 CGAN 生成数据进行筛选，并使用 IF 对交易样本进行异常性建模，将得到的异常性评分加入数据集中，从而增强模型对欺诈行为的区分能力 [7、8]。

Isolation Tree 是一种基于树结构的无监督异常检测方法，其核心思想是：异常样本更容易被随机划分过程“孤立”，即在较少的分割步骤中即可与其他样本区分 [7]。与传统依赖距离或密度度量的方法不同，*Isolation Tree* 采取一种完全不同的方式：通过随机划分特征空间对样本进行分离，从而根据一个样本被隔离所需的步骤多少来判断其是否异常 [7]。

在每棵 *Isolation Tree* 中，对于一个样本 x，其路径长度 $h(x)$ 被定义为：样本 x 从树根节点走到叶子节点（即被孤立）所需的切分次数。通常情况下，异常样本位于稀疏区域，往往在前几次切分中就被单独隔离出去，路径长度 $h(x)$ 较小 [1]。而正常点路径长：它们被大量其他点包围，需要较多次随机切分才能完全隔离，路径长度 $h(x)$ 较大 [7]。由于算法构造了多棵树，因此对每个样本 x 的路径长度会取平均值：

$$E\left[h(x)\right]=\frac{1}{t}\sum_{i=1}^{t}h_i(x) \quad (3)$$

其中 t 是森林中的树数，$h(x)$ 是第 i 棵树中样本 x 的路径长度。

为了便于不同规模下的比较，*iForest* 引入了路径长度的归一化项：调和期望路径长度：对于包含 n 个样本的数据集，树的平均路径长度期望为 [7]：

$$C(n)=2H(n-1)-\frac{2(n-1)}{n} \quad (4)$$

最终异常评分定义为 [7]：

$$S(x,n)=2^{-\frac{E[h(x)]}{C(n)}} \quad (5)$$

当 $E[h(x)]$ 很小（即容易孤立）时，$S(x,n)\to 1$，代表高度异常；

当 $E[h(x)]$ 很大（即难以孤立）时，$S(x,n)\to 0$，代表正常点 [7]。

（三）XGBoost 模型

XGBoost 是一种基于梯度提升（Gradient Boosting）的集成学习模型，其基本学习器为 CART 回归树 [24]。从统计学习角度看，*XGBoost* 通过构建加性模型，对交易特征与欺诈标签之间的复杂非线性关系进行建模，从而实现对条件分布 $P(Y=1|X=x)$ 的有效逼近 [25、26]。

设交易样本定义为 (X,Y)，其中 $x\in R^d$ 表示包含了 d 维特征的交易特征向量，$Y\in\{0,1\}$ 为对应的类别标签，分别表示正常交易与欺诈交易，对于给定的训练样本集 $D=\{(x_i,y_i)\}_{i=1}^{n}$，*XGBoost* 模型通过构建多颗回归树组成的加型模型，对样本的预测输出表示为：

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i), f_k \in F \quad (6)$$

其中 F 表示 CART 回归树构成的函数空间，K 为基学习器的数量。

在模型训练过程中，*XGBoost* 通过最小化经验损失与正则化项之和来学习模型参数，其目标函数定义为：

$$OBj = \sum_{i=1}^{n} L\left(y_i, \widehat{y_i}\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \quad (7)$$

其中 $L(\cdot)$ 为损失函数，用于衡量预测值与真实标签之间的误差，$\Omega(\cdot)$ 为正则化项，用以约束模型复杂度并防止过拟合。在本文中，基学习器采用平方误差损失函数：

$$L\left(y_i, \widehat{y_i}\right) = \left(y_i - \widehat{y_i}\right)^2 \quad (8)$$

对应的正则化项定义为：

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \quad (9)$$

其中 T 表示回归树的叶节点数量，ω_j 为第 j 个叶节点的输出权重，γ 与 λ 为正则化系数，用于控制模型复杂度 [24、25]。

在优化过程中，*XGBoost* 利用损失函数关于预测值的一阶与二阶梯度信息，对目标函数进行二阶泰勒近似，并在每一轮迭代中生成一棵新的回归树，从而在保证计算效率的同时提升模型的稳定性与泛化能力 [24、27]。

在本文提出的 CGAN－IF－Bayes－XGB 框架中，*Isolation Tree* 模型为每一笔交易样本计算异常评分，记为 S_i，该评分用于刻画样本在特征空间中的异常程度。最终，每个交易数据的输入特征向量被扩展表示为：$\tilde{x}_i = [x_i, s_i]$，即将异常评分作为附加统计特征与原始交易特征共同输入 *XGBoost* 分类模型。需要强调的是异常评分仅作为辅助特征参与建模，从而增强模型对潜在欺诈行为的区分能力，而非直接作为分类标签。

（四）使用贝叶斯优化器对 XGBoost 中参数进行优化

在 *XGBoost* 模型中，我们希望找到一组最优超参数 $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$，使得模型在验证集上的性能指标——AUC（Area Under the ROC Curve）达到最优 [10、28]。这个优化问题可表示为：

$$\theta^* = \arg\min_{\theta \in D} f(\theta) \quad (10)$$

其中，$f(\theta)$ 为验证误差（目标函数），通常是无法解析的“黑盒”；D 是超参数空间；$\theta = \{n\_estimators, \max\_depth, min\_child\_weight, learning\_rate, \gamma, \lambda\}$ 表示 *XGBoost* 模型的超参数集合。

从统计建模角度看，贝叶斯优化通过构建目标函数在超参数空间中的概率代理模型，对模型性能的不确定性进行刻画。在给定前 t 次迭代中已观测到的超参数 - 性能对：

$$D_{1:t} = \left\{\left(\theta_i, f(\theta_i)\right)\right\}_{i=1}^{t} \quad (11)$$

贝叶斯优化器通过估计目标函数的后验分布：

$$P\left(f(\theta) \mid \theta, D_{1:t}\right) \quad (12)$$

来预测不同超参数配置下模型性能的潜在表现，并结合采集函数在探索未知区域与利用已有优良参数之间权衡，从而选择下一组最具潜力的超参数进行评估 [28、29]。

本文采用树结构的 Parzen 估计器（Tree-structured Parzen Estimator，TPE）作为代理模型。TPE 通过分别对性能较优与较差的参数区域建立概率模型，在有限计算预算下优先采样潜在最优区域，从而显著提高超参数搜索效率 [11、30]。该方法已被证明在信用评分与欺诈检测等任务中具有良好的稳定性与实用性 [30]。

贝叶斯优化 *XGBoost* 的工作流程如下图 2 所示：

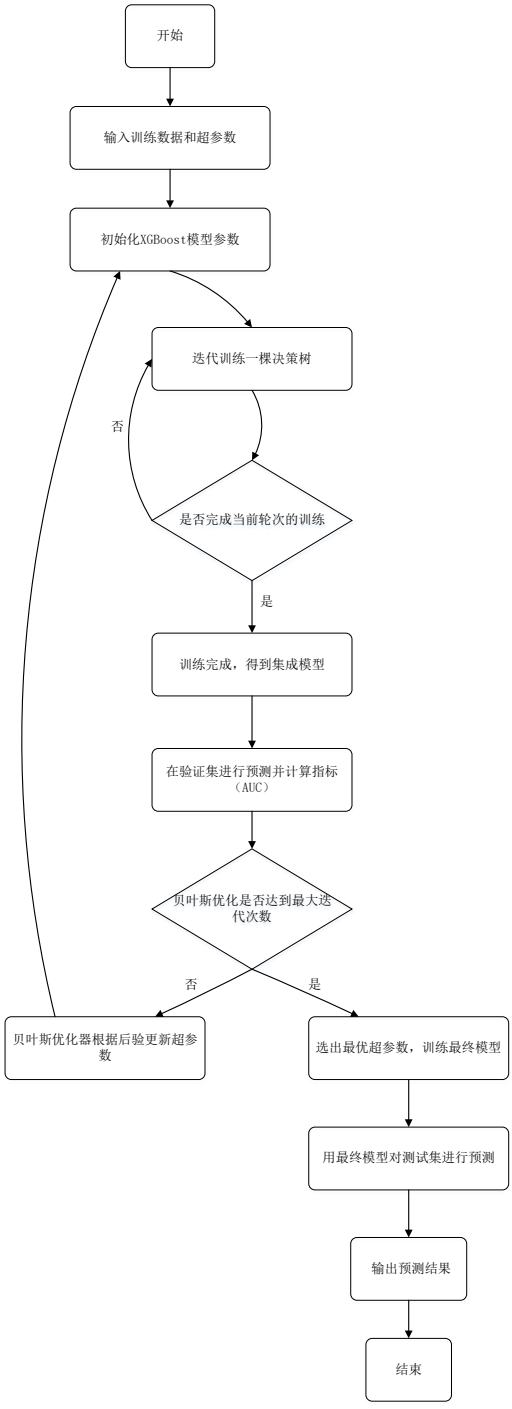

图 2：贝叶斯优化 XGBoost 流程图

三、数据集说明

验证所提出模型在信用卡欺诈检测任务中的有效性与优越性，本文选用 Worldline 公司与比利时布鲁塞尔自由大学（Université Libre de Bruxelles，ULB）机器学习小组在大数据挖掘与欺诈检测研究合作中公开的数据集进行实验分析。该数据集已在 Kaggle 平台公开发布，被广泛用于信用卡欺诈检测相关研究。

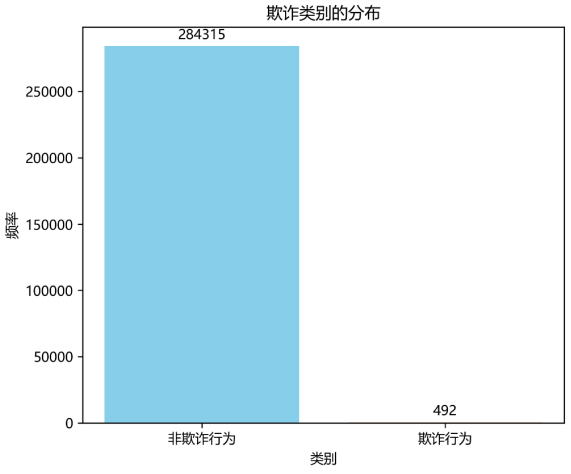

图3：欺诈类数据分布直方图

该数据集包含连续两天内记录的信用卡交易数据，共计284,807笔交易，其中欺诈交易仅492 笔，占比约为 0.17%，具有典型的信用卡欺诈数据特征，数据集类别分布如图3所示。

数据集的特征变量由28个经过主成分分析（Principal Component Analysis，PCA）处理后的匿名特征以及交易金额（Amount）组成。由于交易金额特征与其他经 PCA 处理后的特征在数值尺度上存在显著差异，本文首先对交易金额进行归一化处理，以消除不同特征量级差异对模型训练的影响。

在完成数据预处理后，按照8:2的比例将数据集划分为训练集和测试集，其中训练集用于模型训练与参数优化，测试集用于评估模型的泛化性能[26]。

四、讨论与实验结果

（一）数据集处理

为评估CGAN+IF 方法在数据增强与样本筛选方面的有效性，本文基于训练集对生成结果进行分析。鉴于原始数据集中包含较多特征，为便于可视化比较与分布分析，本文选取具有代表性的关键特征进行实验验证。

在特征选择方面，本文利用 *XGBoost* 模型中提供的 *get_fscore* 函数计算各特征在模型训练过程中的重要性得分[31]，然后将它们的特征分数按从大到小进行排序，如下图4所示，可以看出不同特征对模型预测结果的贡献存在明显差异，其中 V14和 V29的贡献最大。

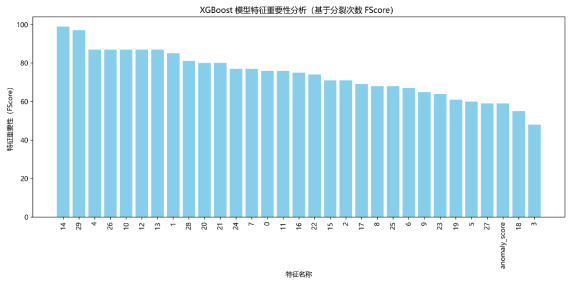

图4：特征重要性排序图

基于上述结果，本文选取特征 V29和 V14这两个特征对 CGAN 生成的欺诈样本与原始欺诈样本的分布情况进行对比分析，如下图5所示。可以观察到，生成的欺诈样本主要分布在原始欺诈样本的高密度区域，表明 CGAN 能够较好地学习真实欺诈数据的潜在分布特征。同时 *Isolation Tree* 模型对生成样本进行了进一步筛选，成功剔除了部分偏离真实分布的异常样本，从而提升了合成数据的整体质量。

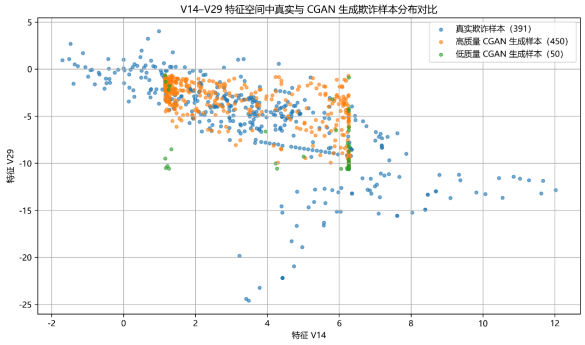

图5：生成欺诈数据与原始欺诈数据分布对比图

此外，为分析 CGAN 的训练稳定性，表1给出了不同训练轮次（Epoch）下判别器与生成器损失函数的变化情况。可以看出，随着训练轮次的增加，判别器在真实样本与生成样本上的损失值整体呈下降趋势并逐渐趋于稳定，说明判别器训练过程较为平稳。生成器损失在训练初期存在一定波动，但随后逐步收敛，表明 CGAN 在训练过程中逐渐学习到真实欺诈样本的分布特征，整体训练过程未出现明显的模式崩溃（Mode Collapse）现象。

表1：CGAN 训练过程中生成器与判别器损失变化

训练轮次（Epoch）	判别器损失（真实样本）	判别器损失（生成样本）	生成器损失
0	0.9634	0.6085	0.8477
500	0.6138	0.6396	0.7517
1000	0.5935	0.6226	0.7883
1500	0.5775	0.5826	0.8333
2000	0.5568	0.5506	0.8905

（二）模型性能评估

1. 贝叶斯优化器与 XGBoost 模型设计

本文将训练集按照8:2的比例进一步划分为训练集与验证集[26]，其中80% 的数据来训练 *XGBoost* 模型，20% 的数据用于模型验证，并采用 AUC（Area Under the ROC Curve）作为性能评价指标，以衡量模型在类别不平衡条件下的分类能力[17]。贝叶

斯优化器以验证集 AUC 作为目标函数，迭代搜索30次，并最终选择 AUC 值最高时对应的 *XGBoost* 超参数组合作为最终模型[10、17]。

由下图6可以看出，随着贝叶斯优化迭代次数的增加，模型性能逐步提升并趋于稳定，最终获得的 *XGBoost* 模型在验证集上的 AUC 值达到0.99以上，表明贝叶斯优化策略能够有效提升 *XGBoost* 模型在信用卡欺诈检测中的判别能力。

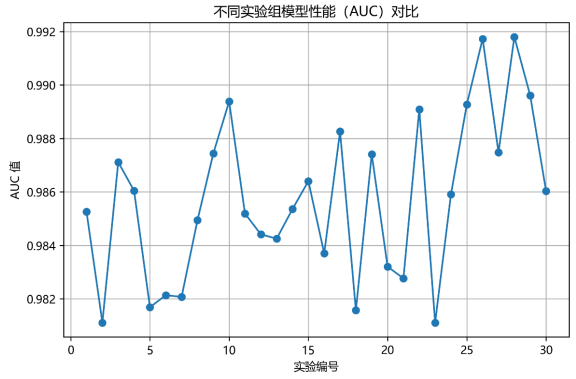

图6：在贝叶斯优化器下 XGBoost 的 ACU 得分

2. 模型评价指标

使用 *XGBoost* 模型对每一笔交易输出一个预测得分（score），该得分可理解为交易属于欺诈类别的置信度或风险水平[32]，因此，在实际应用中需要确定一个合适的分类阈值，以判断当预测得分高于该阈值时是否将交易判定为欺诈[32]，为了获得满足特定性能要求的最优阈值，通常需要对阈值进行微调和移动，并综合评估不同阈值下模型的分类性能[33]，本文通过多种评价指标对不同阈值进行比较分析，最终确定最优决策阈值。相关评价指标包括 MCC、ACC、TPR 和 TNR 等[17、33]。模型性能的评估指标定义如下：

$$TNR = \frac{TN}{TN + FR} \tag{13}$$

$$FPR = \frac{FP}{TN + FP} \tag{14}$$

$$TNR = \frac{TN}{TN + FP} \tag{15}$$

$$TPR = \frac{TP}{TP + FN} \tag{16}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{18}$$

其中 TP 定义了真阳性，表示模型准确预测了欺诈标签。FN 表示假阴性，表示模型未能准确预测欺诈标签。FP 表示误报，表示模型错误地预测了欺诈标签。TN 代表真阴性，表示模型对非欺诈标签的准确预测。TPR 是真阳性率，表示准确预测的欺诈样本的比例。TNR 是真阴性率，表示准确预测的非欺诈样本的比例。FPR 是假阳性率，表示将非欺诈样本预测成欺诈样本的比例。马修斯相关系数（Matthews Correlation Coefficient, MCC）用于衡量模型预测结果与真实标签之间的整体相关性，其取值范围为[-1,1]。其中，1表示完全正确的预测，0表示与随机预测效果相当，-1表示完全错误的预测[33]。

图7展示了在不同阈值条件下 ACC、MCC、TPR 和 TNR 的变化情况。从图中可以看出，当阈值设置为0.4时，各项指标均取得了较为平衡且优良的表现，因此本文将该阈值作为最终模型的判定阈值。

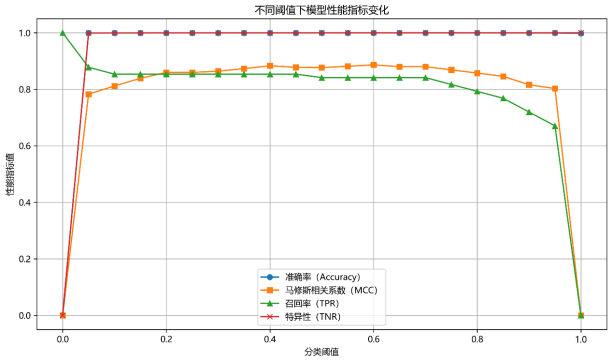

图7：不同阈值下的 ACC、MCC、TPR 和 TNR 得分

3. 不同模型的效果对比

在实验的最后，我们将 CGAN-IF-Bayes-XGB 与相关方法进行性能对比，如下表2所示，我们可以看出综合四个指标 CGAN-IF-Bayes-XGB 模型在对信用卡欺诈识别上有着显著的优势。

表2：CGAN-IF-Bayes-XGB 与对比方法的分类性能比较

模型	准确率	召回率	特异性	马修斯系数
CGAN-IF-Bayes-XGB	0.99960	0.86585	0.99980	0.88333
KNN	0.96914	0.88353	0.97113	0.59032
XGBoost	0.96963	0.85294	0.99954	0.81003
Random Forest	0.95832	0.80253	0.99894	0.69024
AE-based Clustering	0.98902	0.81632	0.98933	0.30585

（三）模型讨论与应用

根据实验结果我们可以得出 CGAN-IF-Bayes-XGB 模型在处理不平衡数据集时有着更大的优势，同时与传统机器学习模型以及深度学习模型相比在处理信用卡欺诈数据集时优势显著，与 XGBoost 相比在具有代表性的 ACC 指标上提升了3.09%，MCC 指标上提升了9.05%，与 KNN 相比在 ACC 指标上提升了3.14%，在 MCC 指标提升了49.64%，这表示在引入 CGAN 和 Isolation Tree 模型后对 XGBoost 的性能有着较大的提升，但模型仍然存在着训练复杂度较高的问题以及 TPR 指标不如 KNN 的现象，未来我们将从优化模型复杂度方面入手，旨在开发出性能更好的 CGAN-IF-Bayes-XGB 模型，让其应用在更广的风控场景。

五、结束语

本文针对银行信用卡交易中类别高度不平衡和欺诈样本信息不足的问题，提出了一种融合数据生成、异常检测与监督分类的 CGAN-IF-Bayes-XGB 欺诈检测方法。该方法通过 CGAN 扩充少数类欺诈样本，并结合 Isolation Forest 提取异常评分作为辅助特征，有效增强了训练数据的判别信息表达；同时，引入贝叶

斯优化对 XGBoost 模型的超参数进行自动调优，并通过阈值搜索确定最优决策边界。实验结果表明，在真实银行匿名交易数据集上，所提出方法在多项评价指标上均显著优于 XGBoost、KNN 等对比模型，尤其在不平衡数据场景下表现出更强的稳定性和综合判别能力。本文的研究结果表明，将条件生成模型、异常检测机制与梯度提升分类器进行有机融合，是缓解信用卡欺诈检测中数据不平衡与特征表达不足问题的一种有效途径。未来研究将进一步关注模型计算复杂度的优化，并在不同来源的银行风控数据集上验证其泛化能力。

参考文献

[1] Nassif., A.B.Talib., M.A.Nasir., Q.Dakalbab., F.M. Machine Learning for Anomaly Detection: A Systematic Review[J]. IEEE Access, 2021, 9, :pp.78658 - 78700.

[2] THE NILSON REPORT. Global Card Fraud Losses at $33 Billion[R]. GlobeNewswire, 2026-01.

[3] 人民网 . 信用卡诈骗案一年5万起 银行损失数百亿仅追回16.5亿元 [N]. 2016-08.

[4] SUBUDHI S., PANIGRAHI S., DASH S.R. ARMS: Automated rules management system for fraud detection[EB/OL]. (2020-02)[2026-01-31]. http://arxiv.org/abs/2002.06075.

[5]LEE J., PARK K. GAN-based imbalanced data intrusion detection system[J]. Personal and Ubiquitous Computing, 2021, 25: 121-128.

[6]GOODFELLOW I.J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y. Generative Adversarial Nets[C]// GHAHRAMANI Z., WELLING M., CORTES C., LAWRENCE N.D., WEINBERGER K.Q., eds. Advances in Neural Information Processing Systems 27. Montreal, Canada: Neural Information Processing Systems Foundation; Curran Associates, Inc., 2014: 2672-2680.

[7] Liu F.T., Ting K.M., Zhou Z. Isolation-based anomaly detection ACM Trans[J]. Knowl. Discov. Data (TKDD), 2012, 6 (1): 1-39.

[8]Li, X., Liang, D., Li, X. et al.. Quality monitoring of real-time PPP service using isolation forest-based residual anomaly detection[J]. GPS Solutions, 2024, 28(3): 118.

[9]LIU F.T., TING K.M., ZHOU Z. Isolation Forest[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(8): 413-422.

[10]SNOEK J., LAROCHELLE H., ADAMS R.P. Practical Bayesian Optimization of Machine Learning Algorithms[C]// LECUN Y., BENJIOU Y., HINTON G., eds. Advances in Neural Information Processing Systems 25. Lake Tahoe, NV, USA: Neural Information Processing Systems Foundation; Curran Associates, Inc., 2012: 2951-2959.

[11]BERGSTRA J., YAMINS D., COX D.D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures[C]// GOMEZ F., TUKEY J., EDWARDS R., eds. Proceedings of the 30th International Conference on Machine Learning (ICML 2013). Atlanta, GA, USA: JMLR.org, 2013: 115-123.

[12] CHEN S., GOO Y.-J.J., SHEN Z.-D. A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements[J]. The Scientific World Journal, 2014, 11: 1-9.

[13] ALAMRI M., YKHLEF M. Survey of Credit Card Anomaly and Fraud Detection Using Sampling Techniques[J]. Electronics, 2022, 11: 4003.

[14]CHOI E., BISWAL S., MALIN B., DUKE J., STEWART W.F., SUN J. Generating multi-label discrete patient records using generative adversarial networks[C]// SMITH J., WANG L., eds. Proceedings of the Machine Learning for Healthcare Conference. Boston: Springer, 2017: 286-305.

[15]FENG X., ZHAO K., ZENG H. Combining GANs with anomaly detection for generating high-quality minority samples[J]. Knowledge-Based Systems, 2020, 193: 105408.

[16] STRELCENIA E., PRAKOONWIT S. A survey on GAN techniques for data augmentation to address imbalanced data issues in credit card fraud detection[J]. Machine Learning and Knowledge Extraction, 2023, 5(1): 304-329.

[17]FAWCETT T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.

[18]SUH S., LEE H., LUKOWICZ P., LEE Y.O. CEGAN: Classification Enhancement Generative Adversarial Networks for unraveling data imbalance problems[J]. Neural Networks, 2021, 133: 69-86.

[19]ZAREAPOOR M., SHAMSOLOMOALI P., YANG J. Oversampling adversarial network for class-imbalanced fault diagnosis[J]. Mechanical Systems and Signal Processing, 2021, 149: 107175.

[20]MULLICK Sankha Subhra S., DATTA Shounak, DAS Swagatam. Generative adversarial minority oversampling[C]//Zhou Jie, Feng Shuiwang, eds. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Republic of Korea: IEEE, 2019: 16951704.

[21]MIRZA M., OSINDERO S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014: 2672-2680.

[22]BA HUNG. Improving detection of credit card fraudulent transactions using generative adversarial networks[J]. arXiv, 2019, 1907.03355.

[23]SOMORJIT L., VERMA M. Variants of generative adversarial networks for credit card fraud detection[C]// KAR NIRMALYA, SAHA ASHIM, DEB SUMAN, eds. Trends in Computational Intelligence, Security and Internet of Things: Third International Conference, ICCISIoT 2020, Proceedings. Cham, Switzerland: Springer, 2020: 133-143.

[24]CHEN T., GUESTRIN C. XGBoost: A scalable tree boosting system[C]// KRISHNAPURAM B., SHAH M., SMOLA A.J., AGGARWAL C.C., SHEN D., RASTOGI R., eds. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA: ACM, 2016: 785-794.

[25] 李航 . 统计学习方法 [M]. 北京：清华大学出版社，2012.

[26]HASTIE T., TIBSHIRANI R., FRIEDMAN J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction[M]. 2nd ed. New York, NY, USA: Springer, 2009.

[27]LIANG W., LUO S., ZHAO G., WU H. Predicting hard rock pillar stability using GBDT, XGBoost, and xgboost algorithms[J]. Mathematics, 2020, 8(765): 1-12.

[28] 王刚 ., 李忠伟 ., 任明明 ., 李鹏 . 机器学习 - 贝叶斯和优化方法 [M].2版 . 北京：机械工业出版社 ,2021.

[29]WU J., CHEN X.Y., ZHANG H., XIONG L.D., LEI H., DENG S.H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization[J]. Journal of Electronic Science and Technology, 2019, 17(1): 26-40.

[30]SUN X., XU W., WANG X. Bayesian Optimization for Hyperparameter Tuning in XGBoost-based Credit Scoring Model[C]// LI H., ZHANG Y., eds. Proceedings of the 2020 International Conference on Artificial Intelligence and Big Data (ICAIBD). Chengdu, China: IEEE, 2020: 393-398.

[31]CHEN T., GUESTRIN C. XGBoost: A scalable tree boosting system[C]// KRISHNAPURAM B., SHAH M., SMOLA A.J., AGGARWAL C.C., SHEN D., RASTOGI R., eds. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, CA, USA: ACM, 2016: 785-794.

[32]FRIEDMAN J.H. Greedy Function Approximation: A Gradient Boosting Machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.

[33]HE H., GARCIA E.A. Learning from imbalanced data[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.

