基于 GAN 与随机森林融合的车险欺诈检测研究

王 娴[1] 崔彩霞[1]
WANG Xian CUI Caixia

摘 要 随着汽车保险行业的快速发展，汽车保险欺诈逐渐增加，这不仅给保险公司带来了一定的经济损失，也在一定程度上影响了广大消费者的权益。传统的二元机器学习方法在车险欺诈检测中被广泛使用，但由于车险欺诈数据中欺诈样本数量与正常模式严重不平衡，这些方法往往无法有效识别欺诈，导致检测结果不佳。针对这一问题，文章提出了一种基于 Wasserstein 条件生成网络（CWGAN-GP）的汽车保险欺诈检测方法。此方法通过生成合理的欺诈模式来重新平衡数据集，以提高检测欺诈模式中的分类器的能力。CWGAN-GP 结合 Wasserstein 的距离和梯度惩罚机制，以防止模型崩溃，同时通过条件生成框架稳定训练和学习不同上下文条件下正常和错误模式的分布。随后，将生成的平衡数据集与随机噪声分类器相结合，以检测汽车保险欺诈。实验结果表明，与传统方法相比，提出的基于 CWGAN-GP 的检测模型在实际车险数据集中表现出更高的欺诈检测准确性和鲁棒性，可以有效降低车险欺诈的财务风险。

关键词 车险欺诈检测；CWGAN-GP；类别不平衡；随机森林

doi：10.3969/j.issn.1672-9528.2025.08.025

0 引言

不平衡数据普遍存在于生活的方方面面，不仅数据分布广泛，而且数据比例不均衡。广义上来说，保险欺诈[1]是一种故意误导或歪曲信息的行为，目的是在保险索赔期间获得不应有的好处。在各种类型的保险欺诈中，由于汽车保险欺诈日益复杂和普遍，汽车保险欺诈对保险公司来说是一项重大挑战。查明汽车保险欺诈的目的是通过分析历史索赔数据来确定普通和欺诈性索赔的特点，以便及时和准确地确定新索赔在到达时是否合法。事实上，这是一个二分类问题。然而，在实际的汽车保险数据集中，欺诈性索赔的数量往往比合法索赔少得多，造成数据分配的严重失衡。这一不平衡使得传统分类方法更倾向于多数类，因此很难有效地确定少数群体。发现汽车保险欺诈，将虚假索赔误作合法索赔可能会给保险公司造成重大的财务损失，将合法索赔误判为欺诈性索赔可能会削弱客户对公司的信任。因此，在发现汽车保险欺诈方面，解决这类不平衡问题已成为一项重大挑战。

针对数据分布不平衡的问题，研究者们提出了多种解决方案，主要有两个方向。一个方向是算法级方法，主要由以 Bagging[2] 为代表的综合学习算法和以 Rescaling[3] 为代表的成本敏感算法组成。集成学习的本质是将几个弱分类器组合在一起，每个分类器彼此独立，将多个分类器比例最高的类别视为评分的最终结果。所有弱分类器都不会相互干扰，从而减少了过拟合的机会。由于误分类成本不同，为了让分类器能够更多地关注误分类成本较高的类的样本，成本敏感算法选择将不同的权重分配给不同误分类成本的类，从而提高一些误分类成本较高的类的准确率。另一种方法是使用数据级重采样技术来平衡数据集。随机采样方法只是从少数样本中随机抽取样本，并将来自多数组的相同数量的样本添加到数据集中，以添加新的、平衡的数据集，但这种方法是重复采样。一些少数模式被过度表达，导致熟练模型严重过度适应。同时，如果对少数样本进行重复噪声采样，分类器的性能将继续下降。合成少数技术（SMOTE）[4] 通过随机插值少数模式及其邻居来解决类不平衡问题。这种方法忽略了周围多数样本的分布，很容易造成新生成的少数样本与周围多数样本重叠的问题。

近年来，生成对抗网络（GAN）因其能够通过对抗性学习生成高质量的少数群体模式而被广泛用于解决阶级平衡问题。Fiore 等人[5] 训练 GAN 创建少数样本，然后将其与原始数据相结合以创建新的训练数据集。实验表明，在新训练集中学习的分类器的排名性能优于原始数据集，从而创建有效的欺诈检测策略。Yang 等人[6] 提出了一种缩

1. 太原师范学院 山西晋中 030619

放不平衡数据集的新方法，即 IDA-GAN，使用差分编码器来研究多个类别的分布，这使得 GAN 的学习过程更加稳定和快速。

尽管基于 GAN 的方法取得了一定进展，但其主要聚焦于生成少数类样本，常忽略数据集整体结构，易导致样本重叠或引入噪声。为此，本文提出改进的基于条件生成对抗网络的 CWGAN-GP 模型，通过结合条件信息和梯度惩罚，生成高质量少数类样本并保留数据分布特性。结合随机森林分类器，本文方法有效解决类别不平衡问题，显著提升欺诈检测性能。

1 相关技术

1.1 生成对抗网络

生成对抗网络（GAN）[7] 是由 Goodfellow 等人于 2014 年提出的深度网络的生成框架，至今仍有学者在原始生成对抗网络的基础上进行改进以得到更好的训练效果以及更多样的生成样本种类，这是所有其他 GAN 变体所基于的基础模型。GAN 主要由生成网络（G）和判别网络（D）两部分组成，GAN 结构图如图 1 所示。

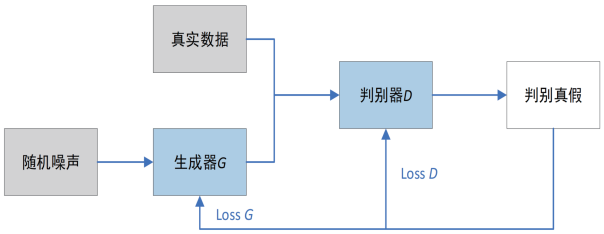

图 1 GAN 结构模型

GAN 的最终目标是使生成网络生成的样本分布与真实数据的分布一致，从而让判别网络无法分辨样本真假，达到一种纳什均衡状态。loss_{GAN} 可以被表示为：

$$\min_G \max_D E_{x\sim P_{\text{data}}}[\log D(x)] + E_{z\sim P_z}[\log(1-D(G(z)))] \quad (1)$$

式中：loss_{GAN} 表示生成器生成的真实样本与假样本的差值；x 表示真实样本数据；z 表示随机噪声， 即经过鉴别器熵后得到的真实样本数据 x 的分布，对应的表示随机噪声 z 的分布经过鉴别器后的熵。

1.2 生成器部分

生成网络 G 通过学习随机噪声 z，尝试生成与原始数据集真实样本分布相近的样本，生成器结构图如图 2 所示。

1.3 判别器部分

判别网络 D 则通过学习真实样本的分布，用以判断生成网络 G 生成的样本是否真实，判别器结果图如图 3 所示。

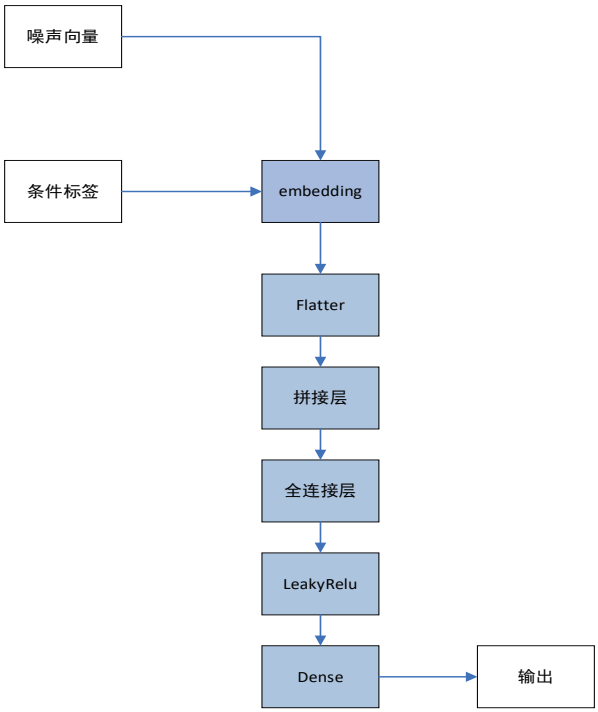

图 2 生成器结构图

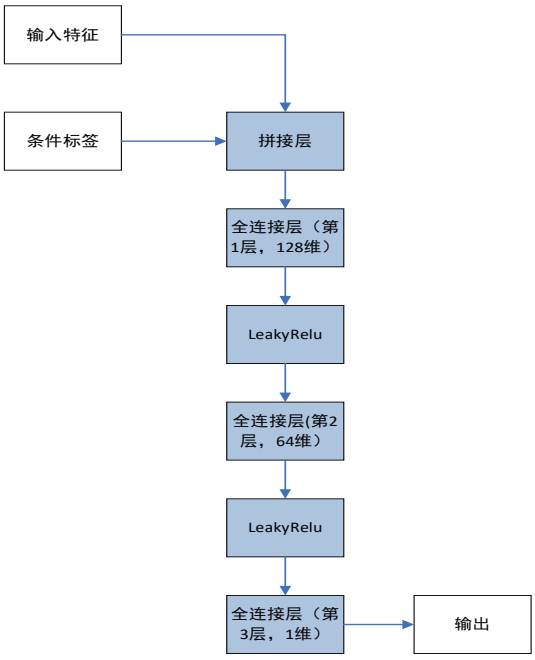

图 3 判别器结构

1.4 CWGAN-GP

CGAN[10] 通过在生成器和判别器中加入条件信息，使得生成过程受到控制，可以生成具有特定属性的样本。通过条件信息的控制，CGAN 能够在生成对抗网络的基础上提供更大的生成灵活性和更高的样本质量。Wasserstein 距离可以用公式表示为：

$$W(P_r,P_g)=\inf_{\gamma\subset\Pi(P_r\sim P_g)} E_{(x,y)\sim\gamma}[\|x-y\|] \quad (2)$$

式中：$W(P_r, P_g)$ 表示衡量两个概率分布 P_r 和 P_g 的距离；inf 表示下确界；$E_{(x,y)\sim\gamma}$表示对联合分布 γ 中的 (x, y) 采样后，计算其期望值；$\|x-y\|$表示 x 与 y 之间的距离。

传统 GAN 使用的是 Jensen-Shannon 散度（JS 散度）或交叉熵损失，当真实数据分布与生成数据分布没有重叠（或部分重叠）时，JS 散度可能失效，导致梯度消失问题，使得生成器无法有效更新。Wasserstein 距离是一种基于线性规划度量两个概率分布之间距离的方法，用于计算将一个概率分布变换成为另一个分布所需要的最小代价。WGAN 引入了 Wasserstein 距离，衡量两个分布之间的“运输成本”。即使生成分布和真实分布几乎不重叠，Wasserstein 距离仍能提供有意义的梯度，避免梯度消失。

梯度惩罚的思想是约束判别器 D 对输入数据的梯度范数接近于 1。具体而言，对于采样点，梯度惩罚项定义为：

$$L_{GP} = \lambda E_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (3)$$

式中：L_{GP} 表示梯度惩罚项的损失函数；λ 表示惩罚项的权重超参数。在 WGAN[9] 中，判别器需要满足 1-Lipschitz 连续性的要求（即判别器的梯度范数不超过 1），是保证 Wasserstein 距离可以被正确估计的必要条件。然而，原始 WGAN 通过简单的权重剪裁强制判别器满足 1-Lipschitz 条件，但是权重剪裁限制了模型的表达能力，导致判别器学习能力不足，还可能引起梯度爆炸或梯度消失。而梯度惩罚通过直接约束梯度范数，更自然且更有效地实现了 1-Lipschitz 条件，还能防止梯度爆炸或梯度消失问题，训练更加稳定。

2 基于 CWGAN-GP 的车险欺诈检测模型

2.1 CWGAN-GP 模型

CWGAN-GP 模型的结构如图 4 所示。CWGAN-GP 模型的生成器、鉴别器和辅助分类器采用全连通层结构。总体结构相似，分为输入层、隐藏层和输出层三部分。隐藏层是 MLP，用来抽象输入数据的特征。为了稳定前向输入分布，加快收敛速度，对每个隐藏层使用层归一化。至于输入和输出层，三者之间存在细微的差异。其中，发生器通过输入符合正态分布的随机噪声 z 和类标号 y 来生成特定样本。发生器在最后一层隐藏层和输出层之间不使用激活函数，在其他层之间使用 leakyrelu 激活函数。鉴别器输入的数据分为两类，一类是生成器生成的新样本数据，另一类是原始训练集中的数据。为避免过拟合，在使用训练集中的原始数据训练鉴别器时，需要在其中加入一定的噪声。鉴别器网络的激活函数用法与发生器相同。辅助分类器结构与鉴别器相似，不同之处在于输入层不再添加类标签，输出层使用 sigmoid 激活函数输出样本预测概率。

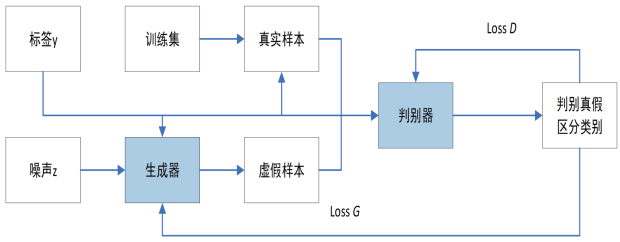

图 4 CWGAN-GP 模型构建

2.2 随机森林

随机森林 [8]（random forest, RF）是由许多决策树组成的模型，通过从预处理后的训练集中有放回抽样，并从所有变量中随机选择变量的方法来生成新的训练集来构造决策树。自 2001 年推出以来，随机森林算法（RF）在分类任务领域取得了巨大成功，已成为一种通用的分类算法。当 RF 用于分类任务时，当有新的样本输入时，RF 会使用多个内部决策树进行判断和分类。每个决策树都会根据以前的学习经验从新样本中接收分类结果。每个决策树的成果不会相互影响。最后，在所有决策树中选择最高百分比的分类结果作为分类的最终结果。随机森林简单易实现，且具有较强的鲁棒性和可解释性，被广泛应用在分类和回归问题汇总，尤其在数据集噪声大、特征维度高、样本不平衡等情况下也有较好表现，与其他分类算法相比，该算法可以快速训练，并且不易过拟合。因此，选择分类器以及上述 CWGAN-GP 模型作为检测汽车保险欺诈的新方法。

3 实验结果与分析

3.1 实验数据

本文采用的实验数据来自 kaggle 公开的车险欺诈检测记录，该数据集名称为 Car Insurance Fraud，含 15 420 条记录，共有 33 个特征。欺诈样本占总样本的 6%，因此该数据集是典型的类不平衡数据集。

3.2 实验对比

为了验证 CWGAN-GP 的有效性，将其与多种现有的数据生成和过采样方法进行了对比，包括传统方法 SMOTE[11]、ADASYN[12] 、B-SMOTE[13] 等，以及生成对抗网络（GAN）和 Wasserstein GAN（WGAN）等先进模型。通过这种对比实验，能够全面评估 CWGAN-GP 在数据生成质量、样本分布特性保持及模型性能提升等方面的优越性，从而验证其在实际应用中的有效性和可靠性。

3.3 评价指标

在二元分类任务中，精度、召回率和 F_1-score 是评估模型预测准确性的关键指标。这些指标来自混淆矩阵，这是一个封装分类算法性能的工具。精度量化了模型正面预

测的准确性。被定义为真正观测值与模型做出的正预测总数的比例。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

召回率是一个度量，用于评估模型在所有真正实例中准确识别正类实例的能力。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

F_1 值代表精度和召回率的调和平均值，将这两个指标整合到一个测量中。

$$F_1 = \frac{2 \bullet \text{precision} \bullet \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

精度、召回率和 F_1 值是衡量分类模型性能的关键工具，从不同的角度帮助理解模型的预测能力。精度关注模型正类预测的可信度，召回率衡量模型能否抓住所有正类样本，而 F_1 值则在两者之间寻找平衡。

AUC（曲线下面积）是指接收操作特征曲线（ROC 曲线）下面积的大小。ROC 曲线的横坐标是假阳性率（FPR），纵坐标是真正率（TPR），AUC 值的范围是 [0,1]，AUC 值越大，表示模型的分类能力越好。

$$\text{AUC} = \int_0^1 (1 - \text{FPR})\, \text{dTPR} \tag{7}$$

式中：TPR 和 FPR 分别为：

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{8}$$

3.4 实验结果

从表 1 中可知，CWGAN-GP 在本实验中表现最优，在 Precision、Recall、F_1 分数和 AUC 上均超过了其他方法，说明该模型可能对数据的特征分布或不平衡问题有更好的处理能力。

表 1 实验结果

模型	Precision	Recall	F_1	AUC
SMOTE	0.884	0.803	0.859	0.895
ADASYN	0.895	0.925	0.886	0.915
B-SMOTE	0.909	0.914	0.878	0.903
CWGAN-GP	**0.932**	**0.928**	**0.912**	**0.932**

由表 2 可知，CWGAN-GP 相比于普通 WGAN 和 GAN，在所有四个指标上都有小幅提升，尤其是 F_1 分数，表明梯度惩罚（GP）和条件生成机制对模型性能有一定的正向贡献。消融实验证明了梯度惩罚和条件生成的结合有效提升了模型的性能。尽管提升幅度不大，但在任务中可能具有实用意义，特别是在精度和召回率都至关重要的情况下。总的来说，这个表证明了 CWGAN-GP 的改进设计在模型性能上优于普通 WGAN，从而验证了改进模块的有效性。

表 2 消融实验

模型	Precision	Recall	F_1	AUC
GAN	0.905	0.869	0.871	0.873
WGAN	0.914	0.921	0.902	0.927
CWGAN-GP	**0.932**	**0.928**	**0.912**	**0.932**

4 总结与展望

在本文中，提出了一种基于 CWGAN-GP 的少数类过采样方法，用于车险欺诈检测。将 CWGAN-GP 方法与现有的过采样算法（如 SMOTE、ADASYN）进行了全面比较。实验结果表明，与其他采样方法相比，CWGAN-GP 在减少假阳性计数的同时能够有效保持真阳性计数的准确性，显著提高了分类性能。这得益于 CWGAN-GP 在少数类数据分布建模上的优势，使得生成的数据更加多样化且具有更高的质量。此外，研究表明，CWGAN-GP 通过优化目标函数（梯度惩罚机制和条件生成），在生成样本质量和分布一致性方面优于传统的 GAN 和 WGAN 模型。消融实验进一步验证了条件生成和梯度惩罚机制对模型性能的重要贡献。由于当前数据集的单一性可能会限制模型的泛化能力和范围，因此未来可以进一步优化此功能。通过引入多种数据源、扩大数据多样性和分布、覆盖案例，尽最大努力打造更具包容性和多样性的数据集。这不仅提高了模型的鲁棒性和准确性，还能使其更好地适应实际应用中的多样化需求，从而克服现有的局限性。

参考文献：

[1] LI Y Q, YAN C, LIU W, et al. A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification[J]. Applied soft computing, 2018, 70: 1000-1009.

[2] YARIYAN P, JANIZADEH S, PHONG T V, et al. Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping[J]. Water resources management, 2020, 34: 3037-3053.

[3] ZHOU Z H, LIU X Y. On multi-class cost-sensitive learning [J]. Computational intelligence, 2022, 26 (3): 232–257.

[4] HAN H, WANG W Y, MAO B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[J]. Advances in intelligent computing,2005,3644: 878-887.

[5] FIORE U, SANTIS A D, PERLA F, et al. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection[J]. Information sciences, 2019, 479: 448-455.

[6] YANG H, ZHOU Y. IDA-GAN: a novel imbalanced data augmentation gan[C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway:IEEE,2021: 8299-8305.

基于 AI 大模型的智能交通系统设计与应用研究

唐 维[1] 杨 超[2*] 顾程琳[3]
TANG Wei YANG Chao GU Chenglin

摘 要 随着人工智能技术的快速发展，AI 大模型在交通领域的应用日益广泛。文章重点探讨 AI 大模型在智能交通系统中的核心技术架构、算法优化、业务应用等关键问题，深入分析大模型驱动的交通智能化解决方案。研究表明，基于 Transformer 架构的大模型在交通流预测、路径优化、事故分析等方面表现出显著优势，预测精度相比传统方法提升 15%~25%。通过构建多模态 AI 融合框架，实现了交通数据的智能处理和实时决策支持，为智慧城市建设提供了重要技术支撑。

关键词 AI 大模型；智能交通；深度学习；交通预测；智能控制

doi：10.3969/j.issn.1672-9528.2025.08.026

0 引言

智能交通系统（intelligent transportation system, ITS）作为现代城市基础设施的重要组成部分，正面临着交通流量激增、城市拥堵加剧、环境污染严重等挑战。传统的交通管理方式已难以满足日益复杂的交通需求，迫切需要引入先进的人工智能技术来提升交通系统的智能化水平 [1]。

1. 浪潮卓数大数据产业发展有限公司 山东济南 250101
2. 中晟软件股份有限公司 山东济南 250014
3. 山东省通信网络保障中心 山东济南 250000

1 AI 大模型驱动的交通系统技术架构

1.1 大模型支撑的交通云脑设计

交通云脑作为智能交通系统的核心中枢，承载着海量交通数据的处理、分析和决策任务。基于 AI 大模型的交通云脑采用多层次架构设计，从底层的数据接入到顶层的用户交互，形成了完整的智能化处理链条。数据接入层负责统一接收来自各类交通传感器、监控设备和第三方数据源的实时信息，通过标准化接口实现异构数据的统一管理 [2]。模型服务层是系统的核心，部署了针对不同交通场景优化的专用大模型，包括基于 Transformer 架构的交通预测大模型、融合图神

[7] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//NIPS'14: Proceedings of the 28th International Conference on Neural Information Processing Systems.NewYork:ACM,2014:2672-2680.
[8] RIGATTI S J. Random forest[J]. Journal of insurance medicine, 2017, 47(1): 31-39.
[9] WENG L L. From GAN to WGAN[EB/OL].(2019-04-18)[2025-05-10].https://doi.org/10.48550/arXiv.1904.08994.
[10] LOEY M, MANOGARAN G, KHALIFA N E M. A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images[J]. Neural computing and applications, 2020(26): 1-13.
[11] FERNÁNDEZ A, GARCIA S, HERRERA F, et al. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary[J]. Journal of artificial intelligence research, 2018, 61: 863-905.
[12] HE H B, 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 world congress on computational intelligence). Piscataway: IEEE, 2008: 1322-1328.
[13] MAJZOUB H A, ELGEDAWY I. AB-SMOTE: an affinitive borderline SMOTE approach for imbalanced data binary classification[J]. International journal of machine learning and computing, 2020, 10(1): 31-37.

【作者简介】

王娴（2001—），女，山西吕梁人，硕士研究生，研究方向：机器学习。

崔彩霞（1974—），女，山西吕梁人，博士，副教授、硕士生导师，研究方向：机器学习、数据挖掘。

（收稿日期：2025-03-05 修回日期：2025-07-29）