

## ◆ 机器学习模型 ◆

融合 GAN 与深度神经网络的混合信用评估方法<sup>\*</sup>

彭柏程,张安勤,张 挺

(上海电力大学计算机科学与技术学院,上海 201306)

**摘要:**随着信用卡和个人贷款业务在金融业的快速增长,如何在信息有限的情况下检测潜在违约或坏账业务已经变得极其重要。信用评分领域面临的主要困难是样本不平衡以及分类器性能不佳,为此本研究首先提出了一种基于表格数据的生成对抗网络 Tab-GAN,从原始数据中生成足够的违约样本;随后设计了一种基于 CNN-LSTM 的混合深度学习模型用于特征提取,该模型包含卷积神经网络(Convolutional Neural Networks, CNN)和长短期记忆神经网络(Long Short-Term Memory, LSTM)两个子模型,分别从用户数据中提取静态局部特征和动态时间特征,并加入时空注意力模块对模型的输出进行重要度计算,从而抽取更关键的信息;最后在分类器层面引入焦点损失函数改进轻量级梯度提升机(Light Gradient Boosting Machine, LightGBM)分类器,实现违约风险的概率输出。在两个真实世界数据集中验证风险预测模型,实验结果表明生成对抗网络可以有效解决样本不平衡问题,CNN-LSTM + LightGBM 模型在各项分类评价指标上均优于信用评分领域的其他先进算法,证明了该模型在信用评分领域的有效性和可移植性。

**关键词:**不平衡数据;信用评估;生成对抗网络;卷积神经网络;长短期记忆神经网络;焦点损失函数

中图分类号:TP520.2060 文献标识码:A 文章编号:1005-9164(2023)01-0121-11

DOI:10.13656/j.cnki.gxkx.20230308.014

随着计算机科学技术的进步,互联网、大数据和数据挖掘技术正带动金融行业的变革,中国消费金融市场持续快速增长,现阶段我国消费金融市场已从 2015 年的 19 万亿元增长到 2020 年的 45 万亿元<sup>[1]</sup>。在行业蓬勃发展的同时,金融行业数据量呈现爆炸性增长,消费信贷需求的增加也带来了更多的金融欺诈事件,这些欺诈行为将严重损害消费者和金融

平台的利益。因此,为尽量减少平台和消费者的损失,研究人员提出了大量模型来预测贷款客户的信用风险水平,并避免违约风险的发生。

信用风险评估本质上可以归结为一种二元分类问题,它预测贷款申请人的违约概率,并据此将贷款划分为违约或非违约<sup>[2]</sup>,进而帮助金融机构审批决策。同时,信用风险评估也是不平衡分类问题。在信

收稿日期:2022-02-22

修回日期:2022-04-16

\* 国家自然科学基金项目(42672114)资助。

## 【第一作者简介】

彭柏程(1997-),男,在读硕士研究生,主要从事信用风险预测、金融大数据挖掘研究,E-mail:747865202@qq.com。

## 【引用本文】

彭柏程,张安勤,张挺.融合 GAN 与深度神经网络的混合信用评估方法[J].广西科学,2023,30(1):121-131.

PENG B C,ZHANG A Q,ZHANG T. A Hybrid Credit Evaluation Method Integrating GAN and Deep Neural Network [J]. Guangxi Sciences, 2023,30(1):121-131.

用评分数据集中违约样本数量远小于正常非违约数, 比如当正常样本与违约样本之比为 99 : 1 时, 由于分类器缺乏少数类样本的输入, 导致结果严重偏向多数类, 而且只需将所有样本划分为正常样本, 模型就能达到 99% 的准确率, 但这样的结果毫无意义, 尽可能识别出违约样本才能使金融机构避免损失<sup>[3]</sup>。因此, 信用风险评估不仅是二元分类问题, 而且需要解决数据不平衡所带来的分类性能下降问题。

鉴于信用评分领域存在样本不平衡问题, 研究人员近年来尝试提高违约样本分类的正确率。目前处理不平衡数据集的主流方法可分为 3 种<sup>[4]</sup>。第 1 种是基于数据层面的预处理, 例如欠采样和过采样方法, 通过消除多数类样本和增加少数类样本来改变不平衡数据的原始比例, 虽然这些方法在某种程度上缓解了数据不平衡的情况, 但是删除多数类样本不可避免地造成信息丢失而使模型无法利用已有的信息, 而且基于少数类样本的数据生成缺乏数据多样性, 一定程度上造成模型的过拟合。第 2 种是基于特征提取层面的方法, 挖掘数据之间的隐藏特征, 其中深度学习模型由于其训练多层神经网络并从大量原始数据中提取高维特征, 被广泛用于特征提取和分类回归任务<sup>[5-7]</sup>。近年来随着深度学习技术研究的不断深化, 许多跨学科研究方法正逐步被引入信用评估领域, 例如用自然语言处理挖掘借贷公司之间的相关性<sup>[8]</sup>, 图神经网络挖掘信贷用户社交网络之间实体的联系<sup>[9-11]</sup>, 自编码器<sup>[12]</sup>运用编码-解码技术实现数据降维等, 上述方法均能实现对数据特征的更深挖掘。然而特征提取算法在处理数据维度匮乏且不平衡的数据集时可能会出现模型难以收敛、过度学习噪声信息等问题, 因此需要足够多的数据进行训练。第 3 种是基于分类算法的改进, 代表方法是集成学习。集成学习通过集成多个弱分类器组成强分类器来避免单个分类器对不平衡数据分类预测造成的误差, 从而提高分类效果<sup>[13]</sup>。如 Chawla 等<sup>[14]</sup>提出在自适应增强模型 (Adaboost) 中引入 SMOTE 算法形成 SMOTE-Boost 算法; Seiffert 等<sup>[15]</sup>在 Adaboost 中引入随机欠采样 (Random Under-Sampling, RUS) 形成 RUS-Boost 算法; Rayhan 等<sup>[16]</sup>采用聚类采样 (Cluster-based Under-sampling) 与集成学习结合的方式降低每次输入数据的不平衡比率, 构建 CUSBoost 算法进行不平衡分类; 袁兴梅等<sup>[17]</sup>提出一种基于代价敏感学习的结构化支持向量机集成分类方法, 提高了不平衡数据的分类精度。集成学习大多数通过堆叠分类

器和分配更多权重给少数类样本来提升性能, 是典型的黑盒模型且难以对结果做出合理的解释, 因此需要结合前两种方法进行优化。

基于上述分析, 本研究从数据、特征提取和算法 3 个层面进行改进和优化, 以提高传统方法的分类效果。在数据层面引入生成对抗网络 (GAN), 与传统的数据生成模型相比, GAN 不需要真实数据就能生成逼近真实数据的合成数据, 可以扩展多样性并避免过拟合。因此, 针对信贷领域存在的数据不平衡问题, 提出一种基于表格数据的生成对抗网络 Tab-GAN, 采用 Wasserstein 距离定义真实数据与生成数据之间的差异, 并在生成器中加入辅助分类器优化损失函数, 同时在网络中加入交叉层计算高维特征交互, 生成足够数量的正样本形成增强数据集。在特征提取和分类算法层面, 根据深度学习方法在信用评分领域的应用, 提出基于深度学习的特征提取模型, 将数据特征分为静态统计特征和动态时序特征, 静态财务数据被输入带残差模块的卷积神经网络, 动态用户行为数据输入到长短期记忆神经网络 (Long Short-Term Memory, LSTM) 以捕捉其时序上的关键信息, 使用注意力模块计算静态统计特征和动态时间特征序列的重要性, 融合两个特征模块, 最后输出至改进的轻量级梯度提升机 (Light Gradient Boosting Machine, LightGBM) 分类器进行预测, 使最终分类结果侧重于提高少数类样本的准确率。

## 1 相关理论

### 1.1 生成对抗网络 (GAN)

GAN 包括两个基本部分, 生成器  $G(\cdot)$  和判别器  $D(\cdot)$ 。生成器的目的是生成假样本, 使判别器误判真实样本以获得高分进而达到生成数据与原始数据分布近乎一致的效果。然而, 原始生成对抗网络生成器的输入是随机噪声数据, 因此生成的数据也是随机混乱的, 不能控制生成的图像或数据属于何种类别, 导致模型容易崩塌。

针对原始 GAN 存在的问题, Mirza 等<sup>[18]</sup>提出了条件生成对抗网络 (CGAN) 用于生成更稳定的数据。在 CGAN 中, 生成器和判别器的输入都加入了条件限制, 模型可以从输入数据中学习标签、类别以及不同模态的数据分布, 使网络能快速收敛且不易崩塌。在该模型中, 符合高斯分布的随机噪声  $z$  以及条件  $c$  构成生成器  $G(\cdot)$  的输入, 随后把真实数据  $x$ 、条件  $c$  以及生成器生成的数据  $G(z | c)$  一并输入到判别器

$D(\cdot)$  中, 判别器判断输入数据是否为真实数据, 并根据判别结果返回判别器与生成器的损失。利用反向传播梯度下降的原理优化生成器与判别器, 直到判别器无法分辨数据的来源时, 模型训练终止。CGAN 模型的目标损失函数如下:

$$\text{Score} = \min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x | c)] + E_{z \sim p_z(z)} [\log(1 - D(G(z | c)))], \quad (1)$$

式(1)中,  $E_{x \sim p_{\text{data}}(x)}$  表示真实数据分布的期望函数,  $E_{z \sim p_z(z)}$  表示合成数据分布的期望函数,  $c$  表示条件,  $V(D, G)$  代表真实数据分布和生成数据分布之间的 JS 散度。模型训练的目的在于希望生成器  $G(\cdot)$  的损失  $\log(1 - D(G(z | c)))$  越小越好, 判别器  $D(\cdot)$  的损失  $\log D(x | c)$  越大越好。

### 1.2 卷积神经网络(CNN)

CNN 是深度学习的代表性算法之一, 它由一系列可学习的卷积核与特征图组成, 每一层特征图通过卷积核与上一层特征图的局部区域链接。这些卷积核在图像数据上滑动以提取特征, 对卷积操作的结果加上偏置项  $b_l$ , 经过一个非线性激活函数  $f$  计算该层的结果, 卷积计算公式如下:

$$X_l = f(W \times X_{l-1} + b_l), \quad (2)$$

其中,  $X_l$  和  $X_{l-1}$  分别表示  $l$  层与  $l-1$  层的神经元信息,  $W$  表示权值参数矩阵,  $b_l$  代表偏置项, 通过卷积操作可以提取上一层的局部特征信息。池化层用于缩减网络中的参数数量和模型计算量, 从而提高模型效率并控制过拟合, CNN 结构如图 1 所示。

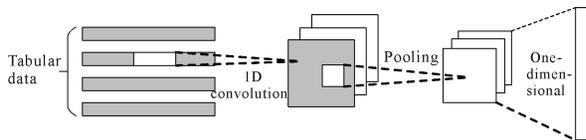


图 1 CNN 结构

Fig. 1 Structure of CNN

### 1.3 长短期记忆神经网络(LSTM)

LSTM 是基于循环神经网络(Recurrent Neural Networks, RNN)的长短期记忆神经网络。这种结构使用特殊的记忆门、遗忘门单元, 人为地控制和调整历史信息的保存比例, 从而允许梯度在网络中传递而不会很快消失。LSTM 结构如图 2 所示。

第一步, LSTM 用遗忘门单元决定被消除的信息。遗忘门单元会读取上一时刻的隐藏状态  $h_{t-1}$  和当前时刻的输入  $x_t$ , 与权重矩阵  $W_f$  相乘加上偏差  $b_f$  后, 使用 sigmoid 函数  $\sigma(\cdot)$  输出一个范围在 0-1 的值  $f_t$ , 继而可以保留网络中的有效信息。

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f). \quad (3)$$

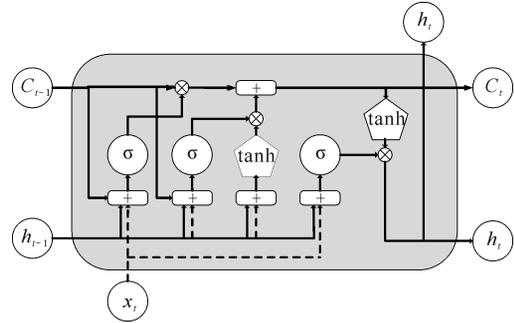


图 2 LSTM 结构

Fig. 2 Structure of LSTM

第二步, 首先使用输入门确定要存储的新信息, 将通过输入门的当前时刻输入  $x_t$  与上一时刻的隐藏状态  $h_{t-1}$  组合, 乘以权重矩阵  $W_i$  并加上偏差项  $b_i$ , 使用 sigmoid 函数  $\sigma(\cdot)$  和 tanh 函数计算候选向量  $\tilde{C}_t$  (其中  $W_c$  代表权重矩阵,  $b_c$  代表偏差项), sigmoid 函数的输出值将决定 tanh 函数的输出值中重要的且需要被保留的信息。

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (4)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c). \quad (5)$$

紧接着, LSTM 单元将遗忘门的输出  $f_t$  与前一层的细胞状态  $C_{t-1}$  相乘, 与输入门的输出  $i_t \times \tilde{C}_t$  逐点相加, 将网络提取的新信息更新到细胞状态  $C_t$ :

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t. \quad (6)$$

最后, 输出门确定输出细胞状态的值, 将前一个隐藏状态  $h_{t-1}$  和当前隐藏状态  $h_t$  传递到 sigmoid 函数中, 然后将新得到的细胞状态  $C_t$  传递给 tanh 函数进行映射, 乘以 sigmoid 函数的输出  $o_t$ , 继而计算出网络下一个节点的隐藏状态  $h_t$ 。

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (7)$$

$$h_t = o_t \times \tanh(C_t). \quad (8)$$

## 2 改进的生成对抗网络

如图 3 所示, 本研究在条件生成对抗网络的基础上优化改进了部分结构, 提出了一种全新的生成对抗网络 Tab-GAN, 其中生成器的输入包含随机噪声  $z$  和条件  $c$ 。由于表格数据具有高维特征, 在输入判别器之前使用交叉层(Crosslayer)计算特征之间的关联性<sup>[19]</sup>, 堆叠  $n$  个交叉层可以提升噪声输入与条件约束之间的交互, 自动构建有限高阶的交叉特征, 从而减少烦琐的人工叉乘。判别器  $D(\cdot)$  中加入一个辅助分类器(Auxiliary Classifier, AC)<sup>[20]</sup>, 使判别器在

输出数据真假的同时输出分类结果  $\hat{c}$ , 同时 AC 的二值交叉熵损失 (Binary cross entropy loss) 被添加到判别器损失中, AC 通过损失函数的反向传播, 使输出分类结果  $\hat{c}$  与生成器输入条件  $c$  相比较, 以鼓励生成器生成特定类别的样本, 如公式(9)所示, 其中  $L_{AC}$  代表预测分类结果  $\hat{c}$  与条件  $c$  的二值交叉熵损失,  $E_{c \sim p_{data}}$  表示条件  $c$  的损失期望函数。

$$L_{AC} = E_{c \sim p_{data}} \left[ -\frac{1}{N} \sum_{i=1}^N c_i \log(\hat{c}_i) + (1 - c_i) \log(1 - \hat{c}_i) \right]. \quad (9)$$

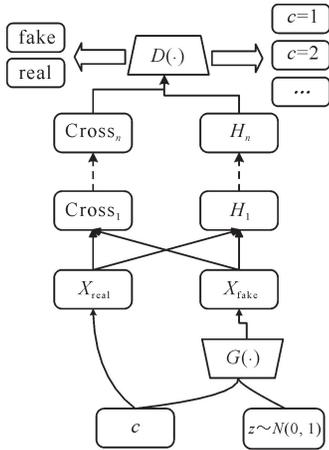


图3 Tab-GAN 结构

Fig. 3 Structure of Tab-GAN

此外, 由于 GAN 一直以来存在难以训练、模式崩塌等问题, 采用 Wasserstein 距离 (简称为“W 距离”) 替换原始生成对抗网络中的 JS 散度<sup>[21]</sup>。不同于 JS 散度与 KL 散度的梯度消失现象, W 距离具有优越的平滑特性, 利用参数数值约束实现生成数据分布与真实数据分布的距离最小化, W 距离公式如下:

$$W(p_g, p_{data}) = \frac{1}{K} \sup_{\|f\| \leq K} E_{x \sim p_{data}(x)}(f(x)) - E_{\hat{x} \sim p_{\hat{x}}}(f(\hat{x})), \quad (10)$$

式(10)中,  $\sup$  表示最小上界,  $\|f\| \leq K$  表示函数  $f$  满足 1-Lipschitz 连续 ( $K$  为 Lipschitz 常数),  $E_{x \sim p_{data}(x)}$  和  $E_{\hat{x} \sim p_{\hat{x}}}$  分别代表真实数据分布与生成数据分布的期望函数, 其中函数  $f$  可以由神经网络拟合, 通过限制网络中所有参数不超过某确信范围使得  $K$  存在。尽管权值剪裁能够使网络符合 1-Lipschitz 条件, 但是由于权值被限制, 并且权值裁剪设置不当会导致网络出现梯度消失的问题。因此, 针对上述问题建立梯度惩罚函数  $L_{GP}$  (Gradient Penalty, GP)。如公式(11)所示,  $E_{\hat{x} \sim p_{\hat{x}}}$  代表生成数据分布的期望函

数,  $\nabla_x D(\hat{x})$  代表判别器梯度, 梯度惩罚项  $\lambda_{GP}$  为软约束, 可以将梯度控制在 1 附近, 缓解梯度消失的问题, 使得模型具有稳定性。

$$L_{GP} = \lambda_{GP} E_{\hat{x} \sim p_{\hat{x}}} \left[ (\|\nabla_x D(\hat{x})\|_2 - 1)^2 \right]. \quad (11)$$

Tab-GAN 训练过程如下。

(a) 创建生成器  $G(\cdot)$  并初始化其权重参数  $\theta_g$ , 创建判别器  $D(\cdot)$  并初始化其参数  $\theta_d$ 。

(b) 生成器输入随机生成的噪声数据  $z = (z_1, z_2, \dots, z_m)$  和条件  $c$ , 输出具有和真实数据分布类似的合成数据集  $X_{fake} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m)$ 。

(c) 计算交叉层  $Cross_n$  与隐藏层  $H_n$  的输出:

$$Cross_n = x_0 x_{n-1}^T \omega_{n-1} + b_{n-1} + x_{n-1}, \quad (12)$$

$$H_n = W_n x_{n-1} + b_n. \quad (13)$$

(d) 将  $Cross_n$  和  $H_n$  进行线性融合后输入判别器  $D(\cdot)$ , 判别器输出数据真假概率  $\hat{y}$  和预测分类  $\hat{c}$ 。

(e) 固定生成器  $G(\cdot)$ , 采用梯度上升的方式更新参数使判别器损失函数  $L_D$  达到最大, 进而提升判别器  $D(\cdot)$  的分类能力。判别器损失函数在原 CGAN 基础上减去梯度惩罚函数  $L_{GP}$ , 并加入辅助分类器的二值交叉熵损失  $L_{AC}$ , 同时使用 Adam 优化器更新判别器参数  $\theta_d$ , 如公式(15)所示,  $E_{x \sim p_{data}(x)}$  代表真实数据分布的期望函数,  $E_{z \sim p_z(z)}$  表示噪声数据分布的期望函数,  $\eta$  代表学习率。

$$L_D = E_{x \sim p_{data}(x)} [D(x)] - E_{z \sim p_z(z)} \cdot$$

$$[D(G(z | c))] - L_{GP} + L_{AC}, \quad (14)$$

$$\theta_d \leftarrow \theta_d + \eta \nabla L_D(\theta_d). \quad (15)$$

(f) 固定判别器  $D(\cdot)$ , 开始训练生成器  $G(\cdot)$ , 生成器损失函数  $L_G$  如公式(16)所示, 采用梯度下降法优化损失函数, 使  $L_G$  达到最小, 从而生成更真实的数据样本, 同时更新生成器参数  $\theta_g$ 。

$$L_G = L_{AC} - E_{z \sim p_z(z)} [D(G(z | c))], \quad (16)$$

$$\theta_g \leftarrow \theta_g + \eta \nabla L_G(\theta_g). \quad (17)$$

(g) 判别器训练 5 次, 生成器训练 1 次, 直到 Tab-GAN 达到纳什均衡。

### 3 基于 CNN-LSTM + LightGBM 的客户分类方法

大多数传统信用评分模型未能利用嵌入数据中的时间依赖关系, 未专注于数据不平衡问题。因此, 为避免数据维度匮乏, 本研究通过 Tab-GAN 模拟生成更多的违约用户样本。通过生成对抗网络生成样本后形成增强数据集, 在此基础上提出一种基于

CNN-LSTM 的混合模型, 该模型包括两个基础学习器 CNN 和 LSTM; CNN 用于提取静态特征, 静态特征主要包含人口特征(年龄、性别、身高体重等)、财务特征(收入、财产等); LSTM 用于提取动态特征, 动态特征包含用户一段时间的操作序列(每个月申请次数、还款次数等)。为捕捉特征提取对象的重要信息并且抑制无关的细节信息, 使用时空注意力模块提取特征向量的重要特征, 并分配一系列注意力权重参数来提升网络性能, 然后用一个全连接层进行融合并输出给 LightGBM 分类器计算分类结果, 模型如图 4 所示。

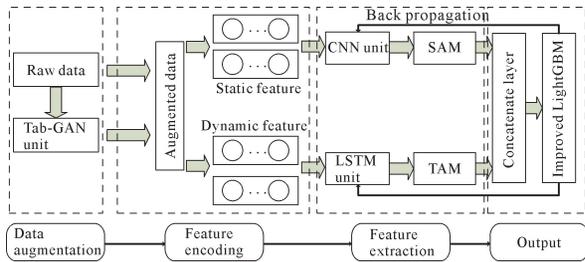


图 4 CNN-LSTM+LightGBM 模型

Fig. 4 CNN-LSTM+LightGBM model

静态特征包含大量低维度表格数据, 可以视作单通道图像。CNN 在处理静态图像类型的数据方面有一定优势, 卷积池化等操作能减少参数的计算量。另外, 本研究的输入表格数据矩阵大小远小于图像, 部署与训练的时间大为减少, 因此使用 CNN 中的卷积层与池化层来提取相应的静态特征。同时, 消费者交易数据包含大量操作行为序列, 可以视为时间序列, 需要序列处理器来获得时间嵌入, 因此 LSTM 用于动态特征的输入。两个特征学习器输出相应的特征向量, 分别加入注意力机制模块以提高模型效果。

注意力机制可以为模型输入序列的每个部分分配不同的权重, 提取更多关键和重要的信息, 使模型能够作出更准确的判断, 同时又不会增加模型的计算和存储成本<sup>[22]</sup>。在信用评分领域已有部分学者使用注意力机制用于特征提取<sup>[23,24]</sup>。然而, 不同类型的注意力机制对数据特征的提取作用存在差异, 信用评分数据集包含多源异构的数据, 需要设计不同的注意力机制模块对其进行重要性计算。本研究采用的信用评分数据集的输入是表格数据以及部分客户操作行为的时间序列, 因此采用混合注意力机制捕捉不同类型数据的重要性信息。表格数据可以看作单通道图像, 引入空间注意力机制进行特征提取, 时间序列则采用时间注意力机制来输出时间点的重要性排序。

空间注意力机制(Spatial Attention Mechanism,

SAM)由图像领域提出<sup>[25]</sup>。如图 5 所示: 首先分别对输入特征图  $X$  做全局最大池化  $\text{MaxPool}(X)$  和全局平均池化  $\text{AvgPool}(X)$ , 得到两个局部特征图; 然后将这两个特征图拼接(concat)生成一个有效的特征表示向量; 随后经过一维卷积运算  $f(\cdot)$ , 再经过 sigmoid 函数  $\sigma(\cdot)$  生成空间注意力特征向量, 即  $M_s(X)$ 。空间注意力机制公式如下:

$$M_s(X) = \sigma(f([\text{AvgPool}(X); \text{MaxPool}(X)])) \quad (18)$$

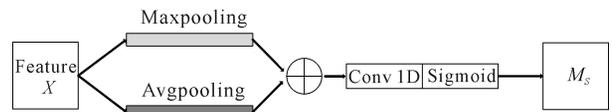


图 5 空间注意力机制

Fig. 5 Spatial attention mechanism

时间注意力机制(Temporal Attention Mechanism, TAM)输入 LSTM 各时刻的隐藏状态, 记忆层将输入向量编码为上下文向量, 注意力层通过调整权重来学习时序数据的内部关系, 并分配注意力给相应时刻。时间注意力机制的计算公式如下:

$$u_i = \tanh(W h_i + b), \quad (19)$$

$$a_i = \text{softmax}(u_i), \quad (20)$$

$$M_T = \sum_{i=1}^t a_i h_i, \quad (21)$$

其中,  $h_i$  代表 LSTM 在  $i$  时间点的输出,  $b$  代表偏差项,  $W$  为权重矩阵,  $u_i$  表示上下文向量,  $t$  表示输入序列的长度,  $a_i$  表示  $i$  时间点输出的权重,  $M_T$  是指每个时间点 LSTM 输出的加权总和。

经过预训练和超参数优化, 对神经网络学习到的两个特征进行线性加权融合, 使用 LightGBM 分类器输出最后的分类结果  $\text{output}_i$ , 违约客户用 1 表示, 正常客户用 0 表示。公式如下:

$$X = M_s \otimes M_T, \quad (22)$$

$$\text{output}_i = F(X), \quad (23)$$

其中,  $\otimes$  代表向量拼接,  $F(X)$  代表 LightGBM 分类器, 使用 Focal loss 损失函数求解和评估模型, 如公式(24)所示。Focal loss 是目标检测算法中为不平衡分类专门设计的损失函数, 其添加了权重因子  $\alpha$  系数和聚焦系数  $\gamma$  使得损失函数更倾向于关注少数类样本, 从而避免易分类样本主导模型训练过程所导致的性能降低。

$$L_{\text{FL}} = \begin{cases} -\alpha(1-\hat{y})^\gamma \log \hat{y}, & y=1 \\ -(1-\alpha)\hat{y}^\gamma \log(1-\hat{y}), & y=0 \end{cases}, \quad (24)$$

其中,  $\hat{y}$  代表分类器的输出概率; 当  $y=1$  时,  $\alpha$  取大

于 0.5 小于 1 的系数可以增大错误分类的损失; 聚焦系数  $\gamma$  能调整易分类样本和难分类样本的损失权重,  $\gamma$  越大则  $(1 - \hat{y})^\gamma$  越小, 以此达到减少易分类样本损失权重的目的, 使分类器更关注难分类样本。

#### 4 数据源与评价指标

本研究采用两个真实世界的客户贷款申请人数据集来验证和评估所提出的模型。第一个数据集取自中国的一家匿名商业银行的用户小额贷款数据, 其中包含大约 15 000 条消费贷款申请记录, 包括资产状况、个人信息、居住城市、使用设备、申请次数等特征, 近 15% 的申请人是违约用户。第二个数据集取自 UCI 机器学习库 (UCI machine learning repository), 它与 30 000 名申请人和交易支付相关, 包含过去 6-12 个月的客户行为数据 (例如每月/季/年的申请量、账单金额、违约历史记录等), 连同他们的财务和人口统计信息 (如性别、工作城市、年龄、财产状况等), 其中有 22% 的申请人是违约用户。将所有数据特征分为静态特征和动态特征, 用 0-1 的标签来表示客户未来是否违约, 数据集样本分布如表 1 所示。

表 1 数据集样本分布

Table 1 Dataset sample distribution

数据集 Dataset	正常样本 Normal samples	违约样本 Default samples	静态特征 Static feature	动态特征 Dynamic feature
Commercial bank data	13 307	2 009	11	17
UCI	23 364	6 636	5	18

随机分离 70% 数据用于训练, 20% 用于测试, 10% 作为验证集。选择信用评分中常用的 4 个指标来评估所提模型的性能: Area Under Curve (AUC)、召回率 (Recall)、F1 值和 Kolmogorov - Smirnov (KS) 值。

True Positive (TP) 表示实际违约且预测也为违约的样本, True Negative (TN) 表示实际未违约且预测也为未违约的样本, False Positive (FP) 表示实际违约但预测为未违约的样本, False Negative (FN) 表示实际未违约但预测为违约的样本。首先, 计算真阳性率 (TPR) 和假阳性率 (FPR) 的值, 公式如下:

$$TPR = \frac{TP}{TP + FN}, \quad (25)$$

$$FPR = \frac{FP}{TN + FP}. \quad (26)$$

然后, 计算准确率 (Accuracy) 来衡量模型分类精度, 并计算召回率和 F1 值以衡量模型对违约样本的识别能力。最后以 FPR 和 TPR 为坐标形成折线图, 即 ROC 曲线。相关值的计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (27)$$

$$Recall = \frac{TP}{TP + FN}, \quad (28)$$

$$F1 = \frac{2TP}{2TP + FP + FN}. \quad (29)$$

KS 是模型中用来区分正负样本分离程度的评价指标, 常用于信用评分模型。每个样本的预测结果是一个 0-1 内的概率值。正负样本的预测概率值由最小到最大排列, KS 值是正负样本分布之间最大差值的绝对值。一般来说, KS 值越大, 正负样本的区分度就越好, KS 值的计算公式如下:

$$KS = \max |TPR - FPR|. \quad (30)$$

CNN-LSTM 神经网络模型的具体参数如表 2 所示, 该网络使用 4 个卷积层, 卷积核参数分别设置为 16, 32, 64, 128 个单元。为防止过拟合, 将随机失活值 (Dropout) 设置为 0.5, 即随机失活 50% 的神经元。采用 ReLU 激活函数和 Adam 优化器加速收敛。

表 2 CNN-LSTM 参数

Table 2 CNN-LSTM parameters

参数 Parameters	值 Value
Batch_size	24/48
Convolution kernel size	16/32/64/128
LSTM units	128
Dropout	0.5
Activation function	ReLU
Optimizer/learning rate	Adam/0.000 1
Epoch	25
Loss function	Focal loss
$\epsilon$	$10^{-8}$

#### 5 仿真实验

为评估所提出 Tab-GAN 的性能, 使用 Python 编程语言, 包含 sklearn、numpy、pandas、tensorflow 和 keras 等工具包进行仿真实验, 实验平台采用 AMD Ryzen 5 3600 3.6 GHz CPU, 32 GB RAM, NVIDIA Geforce RTX 2070 8 G GDDR6 GPU 和

10.2 CUDA 的 64 位计算机对模型进行 8 000 次迭代训练, 并与 CGAN 进行比较。如图 6 所示, Tab-GAN 的生成器损失从一开始就低于 CGAN, 说明添加交叉层可以很好地学习高维数据的交互特征, 提高网络特征提取能力并增强生成器的性能。对于判别器, Tab-GAN 在 7 000 次迭代后的损失明显低于 CGAN, 说明加入梯度惩罚和辅助分类器极大地提高了模型训练的稳定性。对于  $W$  距离, CGAN 在 5 000 次迭代后稳定在 1 左右, Tab-GAN 稳定在 0.6 左右并略有波动, 表明 Tab-GAN 具有更好的收敛稳定性, 并且 Tab-GAN 生成的合成数据分布具有与真实数据更高的相似度。

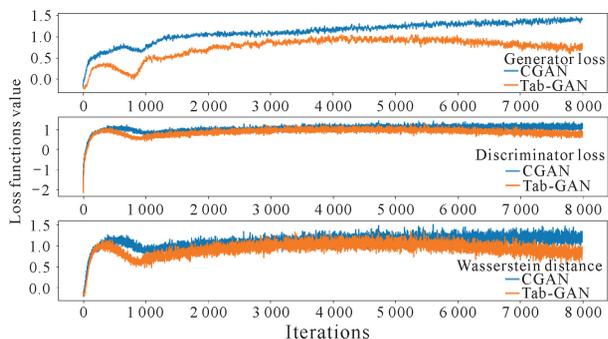


图 6 生成器与判别器损失

Fig. 6 Loss of generator and discriminator

以信贷金额 LIMIT\_BAL、年龄 AGE、4 期账单流水金额 (BILL\_AMT 1-4) 这 6 个特征为例, 绘制真实数据与生成数据的分布频率图。由于数据代入训练模型前经过预处理, 此时特征所代表的值只反映信息, 并不代表真实特征属性值。由图 7 频率密度分布可知, 随着迭代次数的增长, 生成器  $G(\cdot)$  所生成的数据分布愈来愈接近真实数据分布。

为体现本研究数据生成算法的增强效果及数据增强对分类器识别性能的影响, 选用 ADASYN、SMOTE、BorderlineSMOTE、CGAN 等数据增强算法对表 1 中的不平衡数据集进行增强。每个数据集生成一定数量的新样本, 将不平衡比率提高到 0.4, 形成扩充数据集; 采用支持向量机、逻辑回归、决策树、K 近邻 4 种机器学习算法作为对比分类器, 计算 4 种分类器在各项指标上的平均值。表 3 中每列加粗字体表示该项指标的最高值, 从表 3 可以看出, 在利用机器学习算法进行分类时, Tab-GAN 方法在召回率、 $F1$  值、AUC 值和 KS 值上的表现最佳, 总体上明显优于其他 4 种数据增强方法。

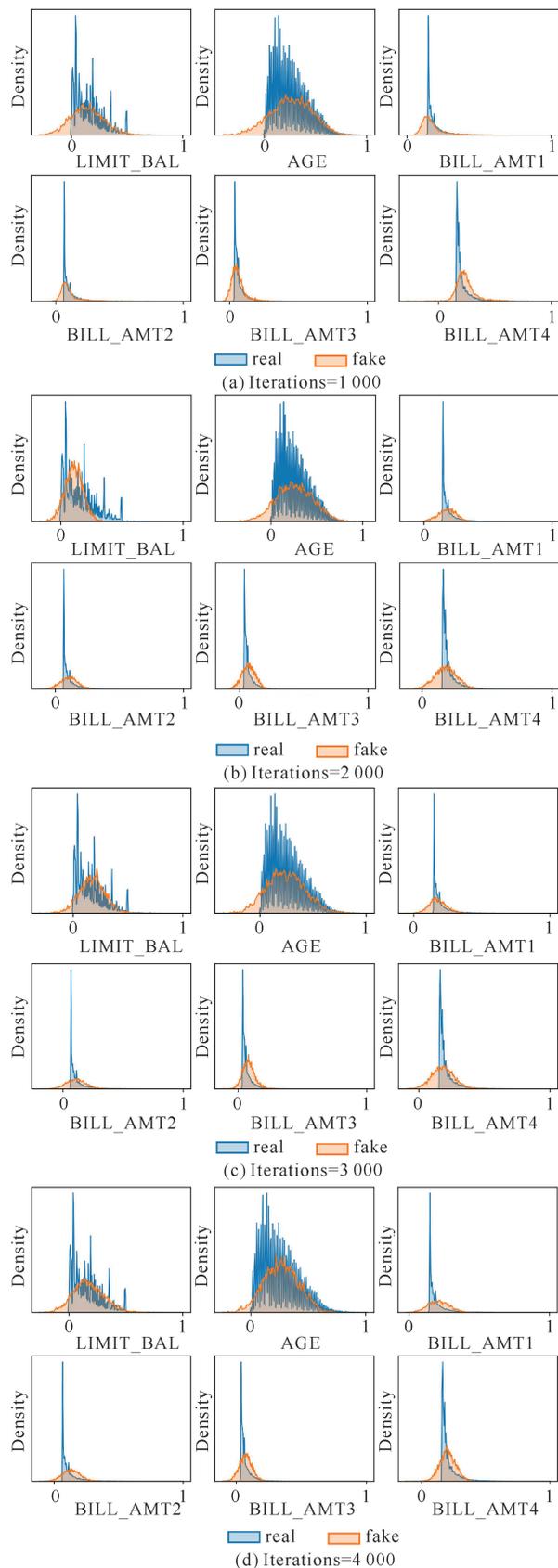


图 7 真实数据与生成数据样本频率密度分布

Fig. 7 Frequency density distribution of real data and generated data samples

表3 数据生成算法性能对比

Table 3 Performance comparison of data generation algorithm

算法 Algorithm	商业银行现金贷数据集 (%) Cash loan dataset of commercialbanks (%)				UCI数据集 (%) Dataset of UCI (%)			
	召回率 Recall	F1	AUC	KS	召回率 Recall	F1	AUC	KS
Base classifier	68.95	56.49	55.50	66.16	61.30	38.07	21.82	30.88
ADASYN	93.27	89.12	82.18	93.49	63.38	43.31	26.19	35.98
SMOTE	92.79	87.91	81.46	92.98	66.83	51.51	32.55	44.37
BorderlineSMOTE	92.98	88.07	81.90	93.43	67.47	52.71	33.90	45.69
CGAN	90.50	81.57	78.97	88.09	68.41	54.42	35.86	46.51
Tab-GAN	<b>93.46</b>	<b>90.05</b>	<b>85.93</b>	<b>94.16</b>	<b>69.02</b>	<b>55.80</b>	<b>37.32</b>	<b>48.38</b>

由于所提模型在特征提取层后加入了注意力层,为此将时间注意力机制与空间注意力机制进行消融实验,结果如表4所示,表4中每列加粗字体表示该项指标的最高值。加入注意力层的模型在4个评估指标上有不同程度的提升,但不同数据集的结果存在差异(表4)。在UCI数据集中,只加入SAM在召回率、F1值、AUC值上的表现优于只加入TAM;而在商业银行现金贷数据集中,只加入TAM在召回率、F1值、AUC值、KS值4项指标均优于只加入SAM,

表4 注意力机制消融实验

Table 4 Ablation experiment for attention mechanism

算法 Algorithm	商业银行现金贷数据集 (%) Cash loan dataset of commercialbanks (%)				UCI数据集 (%) Dataset of UCI (%)			
	召回率 Recall	F1	AUC	KS	召回率 Recall	F1	AUC	KS
Tab-GAN	90.33	81.23	91.45	85.41	55.24	65.52	74.86	52.85
Tab-GAN+SAM	89.35	80.31	90.84	84.38	<b>57.60</b>	66.11	75.37	52.70
Tab-GAN+TAM	94.24	82.22	93.07	84.87	54.51	65.24	74.66	<b>53.58</b>
Tab-GAN+SAM+TAM	<b>96.20</b>	<b>82.52</b>	<b>93.82</b>	<b>85.54</b>	57.32	<b>66.45</b>	<b>75.53</b>	53.49

为证明CNN-LSTM+LightGBM算法的优势,将所提模型与信用评分领域7种先进的分类算法进行比较,包括RNN-RF<sup>[26]</sup>、GAN-AdaBoost-DT<sup>[3]</sup>、CNN、LSTM、CNN-LSTM<sup>[27]</sup>、IGAFN<sup>[28]</sup>、GAN+SAE<sup>[29]</sup>等。实验结果如表5所示,表5中每列加粗字体表示该项指标的最高值。在两个数据集上

表5 模型性能对比

Table 5 Model performance comparison

算法 Algorithm	商业银行现金贷数据集 (%) Cash loan dataset of commercialbanks (%)				UCI数据集 (%) Dataset of UCI (%)			
	召回率 Recall	F1	AUC	KS	召回率 Recall	F1	AUC	KS
RNN-RF	82.07	78.44	87.89	70.64	25.77	35.60	60.24	20.28
GAN-AdaBoost-DT	96.04	78.88	94.24	87.39	49.12	58.93	70.94	41.26

这可能是数据集之间的特征重要性不同导致的。因此,将时间与空间注意力相结合能更加全面提取数据的特征信息,实验结果也表明双注意力机制有效弥补了单注意力机制的性能短板。在UCI数据集中,应用了时空注意力模块的模型在F1值、AUC值上优于其他方法。而在商业银行现金贷数据集中,双注意力机制在召回率、F1值、AUC值、KS值上均优于其他方法,体现出双注意力机制的总体性能最优。

CNN-LSTM+LightGBM模型在召回率、F1值、AUC值、KS值4个指标上相比其他算法具有明显优势,说明本研究提出的CNN-LSTM特征学习算法拥有优秀的特征学习能力,能从不同角度提取嵌入数据节点的信息,同时改进损失函数的LightGBM分类器也能提升不平衡数据学习的分类性能。

续表

Continued table

算法 Algorithm	商业银行现金贷数据集(%) Cash loan dataset of commercialbanks (%)				UCI 数据集(%) Dataset of UCI (%)			
	召回率 Recall	F1	AUC	KS	召回率 Recall	F1	AUC	KS
CNN	80.70	73.38	87.22	88.54	51.02	63.75	73.65	52.85
LSTM	95.09	79.03	93.91	89.46	56.63	66.40	75.44	52.44
CNN-LSTM	98.42	78.81	95.11	87.99	53.34	64.91	74.40	53.29
IGAFN	96.08	81.08	93.87	87.53	50.88	61.01	73.57	43.80
GAN+SAE	84.57	78.98	88.86	81.58	51.64	61.22	72.31	44.49
CNN-LSTM+LightGBM	<b>98.58</b>	<b>82.67</b>	<b>96.90</b>	<b>89.98</b>	<b>57.75</b>	<b>67.96</b>	<b>76.18</b>	<b>54.03</b>

## 6 结束语

在信用评估领域始终存在着数据不平衡问题。本研究引入一种全新的生成对抗网络 Tab-GAN 用于表格数据生成,并对比了其他不平衡数据样本生成算法,Tab-GAN 在模型训练收敛速度以及分类效果提升方面显示出一定的优势。在此基础上提出了一个基于 CNN 和 LSTM 的融合深度神经网络信用评分框架。将特征输入分为静态维度和动态维度,利用 CNN 从静态财务数据中提取静态特征,使用 LSTM 提取动态特征以检测用户操作行为的时间依赖性。随后引入时空注意力模块,使模型在处理表格数据和时序数据时赋予各区域不同的权重,以获取重点关注目标区域的细节信息,并引入 Focal loss 损失函数改进 LightGBM 分类器,提升了神经网络的有监督学习能力,在分类评价指标上具有更好的性能。通过匿名商业银行现金贷数据集和 UCI 数据集上的实验,将混合深度学习模型与相关领域中提出的不平衡分类器的性能进行比较,在召回率、F1 值、AUC 值以及 KS 值评价指标上,本研究所提算法均有明显的提升。

为了在特征提取维度挖掘更多的信息,充分利用平台提供的客户行为数据,未来的工作将关注社交网络数据以进行实体关系的提取,尝试使用图卷积网络来提取嵌入在数据节点中的信息。

### 参考文献

- [1] 王正.消费金融期限结构对我国消费市场规模与增速的影响机制[J].商业经济研究,2021(15):53-56.
- [2] MURALI S,SWAPNA T R. An empirical evaluation of temporal convolutional network for offensive text classification [J]. International Journal of Innovative Technol-

ogy and Exploring Engineering (IJITEE),2019,8(8):2179-2184.

- [3] 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.
- [4] 宋玲玲,王时绘,杨超,等.改进的 XGBoost 在不平衡数据处理中的应用研究[J].计算机科学,2020,47(6):98-103.
- [5] 温小霓,韩鑫蕊.我国科技型中小企业信用风险评价模型——基于 MLP 神经网络的实证分析[J].科技和产业,2017,17(12):159-166.
- [6] 闫静.基于卷积神经网络模型的 P2P 网贷信用风险识别方案策划[D].上海:上海师范大学,2019.
- [7] 徐荣贞,何梦珂.基于 DBN 的线上供应链金融信用风险研究[J].会计之友,2021(11):61-67.
- [8] YAN C,FU X,WU W, et al. Neural network based relation extraction of enterprises in credit risk management [C]//2019 IEEE International Conference on Big Data and Smart Computing (BigComp). Piscataway, NJ: IEEE,2019. DOI:10.1109/BIGCOMP.2019.8679499.
- [9] LIU Z,DOU Y,YU P S, et al. Alleviating the inconsistency problem of applying graph neural network to fraud detection [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM,2020:1569-1572. DOI:10.1145/3397271.3401253.
- [10] HU B,ZHANG Z,SHI C, et al. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, Menlo Park, CA: AAAI,2019,33(1):946-953. DOI:10.1609/aaai.v33i01.3301946.
- [11] ZHONG Q,LIU Y,AO X, et al. Financial defaulter detection on online credit payment via multi-view attribu-

- ted heterogeneous information network [C]//Proceedings of the Web Conference 2020. New York: ACM, 2020:785-795. DOI:10.1145/3366423.3380159.
- [12] AL-SHABI M A. Credit card fraud detection using autoencoder model in unbalanced datasets [J]. Journal of Advances in Mathematics and Computer Science, 2019, 33(5):1-16.
- [13] 曹雅茜, 黄海燕. 基于概率采样和集成学习的不平衡数据分类算法[J]. 计算机科学, 2019, 46(5):203-208.
- [14] CHAWLA N V, LAZAREVIC A, HALL L O, et al. SMOTEBoost: improving prediction of the minority class in boosting [C]//LAVRA N, GAMBERGER D, TODOROVSKI L, et al. Knowledge Discovery in Databases: PKDD 2003. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, 2838:107-119.
- [15] SEIFFERT C, KHOSHGOFTAAR T M, VAN HULSE J, et al. RUSBoost: a hybrid approach to alleviating class imbalance [J]. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 2010, 40(1):185-197. DOI:10.1109/TSMCA.2009.2029559.
- [16] RAYHAN F, AHMED S, MAHBUB A, et al. CUSBoost: cluster-based under-sampling with boosting for imbalanced classification [C]//2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS). Cusboost: cluster-based under-sampling with boosting for imbalanced classification. Piscataway, NJ: IEEE, 2017: 1-5. DOI:10.1109/CSITSS.2017.8447534.
- [17] 袁兴梅, 杨明, 杨杨. 一种面向不平衡数据的结构化SVM集成分类器[J]. 模式识别与人工智能, 2013, 26(3):315-320.
- [18] MIRZA M, OSINDERO S. Conditional generative adversarial nets [Z/OL]. (2014-11-04)[2022-02-01]. <https://doi.org/10.48550/arXiv.1411.1784>.
- [19] WANG R, FU B, FU G, et al. Deep & cross network for Ad click predictions [C]//Proceedings of the ADKDD'17. New York: ACM, 2017:1-7.
- [20] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs [C]//ICML'17: Proceedings of the 34th International Conference on Machine Learning. Sydney: JMLR. org, 2017: 2642 - 2651.
- [21] ZHENG M, LI T, ZHU R, et al. Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification [J]. Information Sciences, 2020, 512:1009-1023.
- [22] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention [C]//NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [23] WANG C, HAN D, LIU Q, et al. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM [J]. IEEE Access, 2019, 7:2161-2168. DOI:10.1109/ACCESS.2018.2887138.
- [24] 王重仁, 王雯, 余杰, 等. 融合深度神经网络的个人信用评估方法[J]. 计算机工程, 2020, 46(10):308-314.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//ECCV 2018: Computer Vision – ECCV 2018. Lecture Notes in Computer Science. [S. l.]: ECCV 2018, 2018:3-19.
- [26] HSU T C, LIOU S T, WANG Y P, et al. Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction [C]//ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2019:1572-1576. DOI:10.1109/ICASSP.2019.8682212.
- [27] NGUYEN T T, TAHIR H, ABDELRAZEK M, et al. Deep learning methods for credit card fraud detection [Z/OL]. (2020-12-07)[2022-02-18]. <https://doi.org/10.48550/arXiv.2012.03754>.
- [28] LEI K, XIE Y, ZHONG S, et al. Generative adversarial fusion network for class imbalance credit scoring [J]. Neural Computing and Applications, 2020, 32(12): 8451-8462.
- [29] CHEN J, SHEN Y, ALI R. Credit card fraud detection using sparse autoencoder and generative adversarial network [C]//2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Piscataway, NJ: IEEE, 2018: 1054-1059. DOI:10.1109/IEMCON.2018.8614815.

# A Hybrid Credit Evaluation Method Integrating GAN and Deep Neural Network

PENG Baicheng, ZHANG Anqin, ZHANG Ting

(College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, 201306, China)

**Abstract:** With the rapid growth of credit card and personal loan business in the financial industry, how to detect potential default or bad debt business with limited information has become extremely important. The main difficulties in the field of credit scoring are sample imbalance and poor classifier performance. For this reason, this study first proposes a generative adversarial network TabGAN based on tabular data to generate sufficient default samples from the original data. Then, a hybrid deep learning model based on CNN-LSTM is designed for feature extraction. The model includes two sub-models: Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). Static local features and dynamic time features are extracted from user data respectively, and the spatio-temporal attention module is added to calculate the importance of the output of the model, thereby extracting more critical information. Finally, the focal loss function is introduced at the classifier level to improve Light Gradient Boosting Machine (LightGBM) classifier and realize the probability output of default risk. The risk prediction model is validated on two real-world datasets, and the experimental results show that the generative adversarial network can effectively solve the problem of sample imbalance. And the CNN-LSTM + LightGBM model is superior to other advanced algorithms in the field of credit scoring in all kinds of classification evaluation index, which proves the effectiveness and portability of the model in the field of credit scoring.

**Key words:** imbalanced data; credit scoring; GAN; CNN; LSTM; focal loss

责任编辑:米慧芝



微信公众号投稿更便捷

联系电话:0771-2503923

邮箱:gxkx@gxas.cn

投稿系统网址:<http://gxkx.ijournal.cn/gxkx/ch>