计算机科学
COMPUTER SCIENCE

可解释的信用风险评估模型:基于注意力机制的规则提取方法

王宝财, 吴国伟

引用本文

王宝财, 吴国伟. 可解释的信用风险评估模型:基于注意力机制的规则提取方法[J]. 计算机科学, 2025, 52(10): 50-59.

WANG Baocai, WU Guowei. Interpretable Credit Risk Assessment Model:Rule Extraction Approach Based on AttentionMechanism [J]. Computer Science, 2025, 52(10): 50-59.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250300059

可解释的信用风险评估模型:基于注意力机制的规则提取方法

王宝财　吴国伟
大连理工大学软件学院　辽宁 大连 116000
(wangbaocai.dlut@163.com)

摘　要　信用风险评估旨在预判客户是否会违约,被视为一项复杂的非线性二分类难题。尽管传统的统计模型在信用评估领域具有一定的应用价值,但其局限性也日益显现。鉴于此,机器学习技术,特别是支持向量机、深度神经网络和集成学习等先进方法,在信用风险评估领域得到了广泛应用,旨在提升模型的准确性和预测精度。然而,尽管这些机器学习模型性能卓越,但其内在的复杂性和不透明性导致模型预测结果难以向用户阐释,在实施过程中面临诸多挑战。为解决这一问题,提出了一种可解释的信用风险评估模型,该模型融合了注意力机制与树集成规则提取技术,能够自动识别训练数据中的复杂非线性关系,实现模型自身的可解释。首先从训练好的树集成模型中提炼出众多可解释的规则,并将这些规则转换为新的特征变量,然后将这些新的特征变量作为注意力神经网络的输入,以精确计算每条规则的注意力权重。在此基础上,模型根据注意力权重、目标函数及约束条件,综合考虑规则子集的预测精度、稳定性和可解释性,可在线性时间内高效地求得最优规则子集。在3个公开数据集上进行了实验,结果表明,所提方法在保持模型较高预测精度的前提下,实现了模型可解释性的显著提升。

关键词:机器学习可解释性;信用风险评估;注意力机制;规则生成算法;树集成模型
中图分类号　TP391

Interpretable Credit Risk Assessment Model:Rule Extraction Approach Based on Attention Mechanism

WANG Baocai and WU Guowei
School of Software Technology,Dalian University of Technology,Dalian,Liaoning 116000,China

Abstract　Due to the limitations of traditional statistical models for credit risk assessment,machine learning techniques have significantly enhanced model accuracy and predictive capabilities. However,the complexity and opacity pose significant challenges in terms of interpretation. To address this issue,this paper introduces an interpretable machine learning model for credit risk assessment that integrates the attention mechanism with tree ensemble rule extraction approach. This model automatically identifies complex nonlinear relationships within the data,extracts a large number of interpretable rules from the trained tree ensemble model,encodes these rules into new feature variables,and inputs them into an attention neural network to obtain attention weights for each rule. Subsequently,based on the attention weights,objective function,and constraints,the model balances the predictive performance,stability,and interpretability of the rule subset. The optimal rule subset can be derived in $O(n)$ time. Experimental results,based on three public datasets,demonstrate that the proposed approach not only maintains high predictive accuracy but also substantially enhances the model's interpretability.

Keywords　Interpretable machine learning,Credit risk assessment,Attention mechanism,Rule generation algorithm,Tree ensemble models

1　引言

随着金融行业的不断发展和全球金融市场的日益复杂,金融机构在贷款审批和信用风险评估方面面临着重大挑战[1]。信用风险评估作为金融机构的核心业务之一,直接关系到投资者的资金安全和市场稳定。传统的信用风险评估统计方法包括线性判别分析[2]、逻辑回归[3-4]、评分卡[5]和贝叶斯分类[6-8],这些方法能在一定程度上提供可靠的预测,但在处理复杂、非线性数据集时,其预测能力和适应性存在不足,而决策树[9-10]、支持向量机[11-13]、人工神经网络[14-16]和集成模型[17-19]等机器学习模型因其强大的预测能力和鲁棒性,在信用风险评估中受到了广泛关注。但这些模型的不透明性和可解释性缺陷已成为阻碍其在实际场景中更广泛应用的关键因素[20]。次贷危机之后,金融机构和监管机构对信用风险评估模型的透明度和可解释性期望提高[21-23]。例如,美国的《平等信用机会法》要求金融机构必须向借款人明确说明拒绝贷款的具体原因[24],欧盟的《通用数据保护条例》(GDPR)则强调使用自动化决策系统时,用户有权获得清晰的解释[25]。这些法律法规进一步凸显了提高信用风险评估模型可解释性的重要性。

到稿日期:2025-03-11　返修日期:2025-05-19
通信作者:吴国伟(wgwdut@dlut.edu.cn)

为了平衡预测准确性和可解释性，研究人员深入研究了规则提取方法[20,26-27]。这些方法旨在用更简单、透明的“白盒”模型（如基于规则的模型或决策树）来复制复杂、不透明的“黑盒”模型（如树集成模型）。这种转换为决策者提供了易于理解的解释和有价值的见解。根据 Martens 等[28]的研究，规则提取方法可分为教学式方法和分解式方法。教学式方法主要关注解释黑盒模型的预测结果，忽略了其内部决策过程的复杂性，导致解释可能与原始模型的准确性不一致。而分解式方法深入探究不透明模型的内部结构，将其转换为可解释的规则，使决策者能够更深入地理解模型的内部机制。

Haddouchi 等[29]提出了 Forest-ORE 框架，通过混合整数优化算法从随机森林中提取优化规则集合。Birbil 等[30]通过数学规划方法最小化规则数量和总不纯度，提高规则的解释性。Manzali 等[31]提出基于关联规则度量的新型森林修剪策 PRM，用于优化决策森林中的分支数量，以在减小模型复杂度的同时提高分类性能。Boruah 等[32]提出透明规则生成随机森林（TRG-RF），通过树排序和规则剪枝提取重要决策规则。Bologna[33]提出一种将随机森林和梯度提升树转换为可解释的多层感知器的规则提取技术。Edali[34]将集合划分公式用于从随机森林中提取规则，以有效减少规则数量。Chen 等[35]基于多目标优化的规则提取方法 RE-MOBDE，从树集成模型中提取可解释规则，用于败血症的早期预测。Shams 等[36]提出 REM 方法，通过从深度神经网络和非深度学习模型中提取规则用于临床决策支持。Wang 等[37]结合多目标优化算法提出一种改进的随机森林规则提取方法 IR-FRE，用于乳腺癌的诊断。Mashayekhi 等[26,38]提出一种基于启发式算法的规则提取方法。Deng[39]广泛探讨了从集成树中提取、评估、简化和选择规则的方法，最终构建了一个稳健的基于规则的学习系统。Dong 等[40]提出一种优化策略，该策略结合了局部和全局规则提取技术，以提高贷款评估模型的可解释性。Friedman 等[41]提出的 RuleFit 算法将提取的规则作为非线性组件集成到线性回归模型中，构建了一个规则集成分类器。Dumitrescu 等[42]基于规则集成的概念，利用简洁的规则制定了一个惩罚性逻辑树回归模型，该模型在信用评分任务中的表现优于传统逻辑回归。Kato 等[43]提出了 SafeRuleFit 算法，该算法通过元安全筛选机制学习最优稀疏规则模型，在保证模型性能的同时增强了规则的可解释性和安全性，其方法在稀疏规则生成方面展现出显著优势。

然而，当前的研究存在以下问题。1）规则提取效率低：基于数学规划的规则提取方法在求解最优规则子集时面临 NP 难问题，难以适用于金融领域的大规模数据集（样本数大约为 10 万）[29-30]。2）指标单一性：现有方法仅优化单一指标（如 AUC）[31-43]，未考虑信用风险评估场景的特殊需求，信贷数据集中“非违约”样本与“违约”样本的比例严重失衡，违约样本作为关键少数类，其识别效果直接影响金融机构的风险控制能力。金融机构不仅关注信用风险评估模型的整体准确率，更强调对违约样本的高召回率，违约漏检会对金融机构产生重大的资金损失。

为了解决当前研究面临的问题，本文首次将随机森林模型的规则提取与注意力机制相结合（称为 RRFA），用于信用风险评估，突破传统方法（如混合整数规划[29]）无法处理大规模数据的限制，创新性地提出考虑多个预测性能指标约束的规则提取方法，并将其应用于金融信用风险评估领域，能够满足信用风险评估模型的可解释性和预测性能需求。首先，在数据集上训练随机森林模型，以捕捉训练数据中的复杂非线性关系。然后，从训练好的随机森林模型中提取大量可解释的规则。接着，将这些规则编码为新的特征变量，并输入注意力神经网络中，获得每条规则的注意力权重。最后，基于注意力权重、目标函数和约束条件，在线性时间内获得最优规则子集，同时综合考虑模型的预测性能、稳定性和可解释性。通过调整约束条件，可以获得满足不同需求的规则子集。在 3 个公共数据集上进行了实验，结果验证了所提方法的有效性和实用性，表明其在保持稳健预测性能的同时，能够显著提升模型的可解释性。

2 背景及相关工作

2.1 问题定义

在信用风险评估的背景下，用户信息由一系列样本对$(\boldsymbol{x}, y)$组成，其中 $\boldsymbol{x}$ 代表用户的多方面综合信息，包括人口统计属性、个人信用历史和财务状况，而 y 表示相应的真实标签，反映用户的信用状况：0 表示信用良好，预期不会违约；1 表示信用不良，存在违约风险。本质上，这是一个二分类任务，旨在开发一个预测模型，即分类器 $H:\boldsymbol{X}\rightarrow\boldsymbol{Y}$，将用户信息 $\boldsymbol{x}$ 映射到其信用类别 y。在这个框架中，$\boldsymbol{x}$ 被形式化为一个 T 维向量$(x_1,\cdots,x_T)$，属于特征空间 $\boldsymbol{X}=X_1\times\cdots\times X_T\subseteq R^T$，而 y 被限制在集合 $\boldsymbol{Y}=\{0,1\}$ 中。信用风险评估模型 H 通过训练和学习来处理包含大量用户样本的数据集，从而准确预测用户的违约概率，辅助贷款决策。通常，H 表现为一个非线性函数，能够捕捉输入特征与用户信用状态之间复杂且非线性的关系。对于更广泛的分类需求，如多类分类，本文概述的基本原理和方法可以相应地进行适应和扩展。

2.2 规则的定义

本文将包含用户信息的训练集表示为 $\boldsymbol{D}=\{(\boldsymbol{x}_i,y_i)\mid i=1,\cdots,N\}$，$\boldsymbol{x}_i$ 表示用户 i 的特征向量，y_i 表示对应的标签。从数据集 $\boldsymbol{D}$ 中导出的分类规则 r_j 表述为$\{c_j=\!>Y_k\}$，c_j 表示规则 r_j 的条件，由变量属性及其值的组合构成，例如$(X_1=small\ \wedge X_2=green)$。规则 $\{c_j=\!>Y=Y_k\}$ 的支持度对应数据集 D 中同时满足条件c_j且结果为 $Y=Y_k$ 的数据条目的比例。条件 c_j 的支持度是数据集 $\boldsymbol{D}$ 中满足c_j的数据条目的比例。规则的置信度计算为该规则的支持度与其条件支持度的比值。

基于信息增益，决策树的非叶节点可以根据属性X^*被划分成 M 个子节点（$M\geqslant 2$）。本文定义 $p(y_k\mid P_v)$为在节点 v 处属于类别y_k的实例比例。节点 v 处 Y 的信息熵如式(1)所示：

$$H(Y\mid P_v)=-\sum_k p(y_k\mid P_v)\log_2 p(y_k\mid P_v) \tag{1}$$

其中，P_v 表示从根节点到节点 v 的属性及其对应值的组合。例如，$P_v=\{X_1=Male\wedge X_2=University\wedge X_5>10\}$。如果 v 是非叶节点，则根据最大化信息增益的标准选择分裂变量 X^*，如式(2)所示：

$$Gain(Y\mid P_v,X^*)=H(Y\mid P_v)-\sum_m w_m H(Y\mid P_v,X^*=a_m) \tag{2}$$

其中，w_m表示子节点v_m中的实例数量占节点v中总实例数量的比例。如果v是叶节点，则采用叶节点中出现最频繁的类别Y_k作为分类结果，并制定一条分类规则为$P_v \Rightarrow Y_k$。决策树中的规则是互斥的，并且没有预定义的顺序。图1给出了一棵决策树，表1列出了从这棵决策树中派生出的规则。

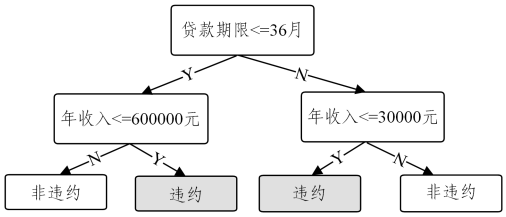

图1 二分类决策树

Fig. 1 Binary classification decision tree

表1 由决策树生成的规则集

Table 1 Rule set generated by the decision tree

规则编号	规则条件	规则结果
r_1	贷款期限≤36月 & 年收入≤600000元	非违约
r_2	贷款期限≤36月 & 年收入>600000元	违约
r_3	贷款期限>36月 & 年收入≤30000元	违约
r_4	贷款期限>36月 & 年收入>30000元	非违约

3 本文模型的设计

本章深入探讨了可解释信用风险评估模型RRFA的整体框架和细节。图2给出了该模型的主要结构，包括5个不同的层：随机森林规则生成层、规则编码层、注意力网络层、用户信用评估层和规则子集提取层。

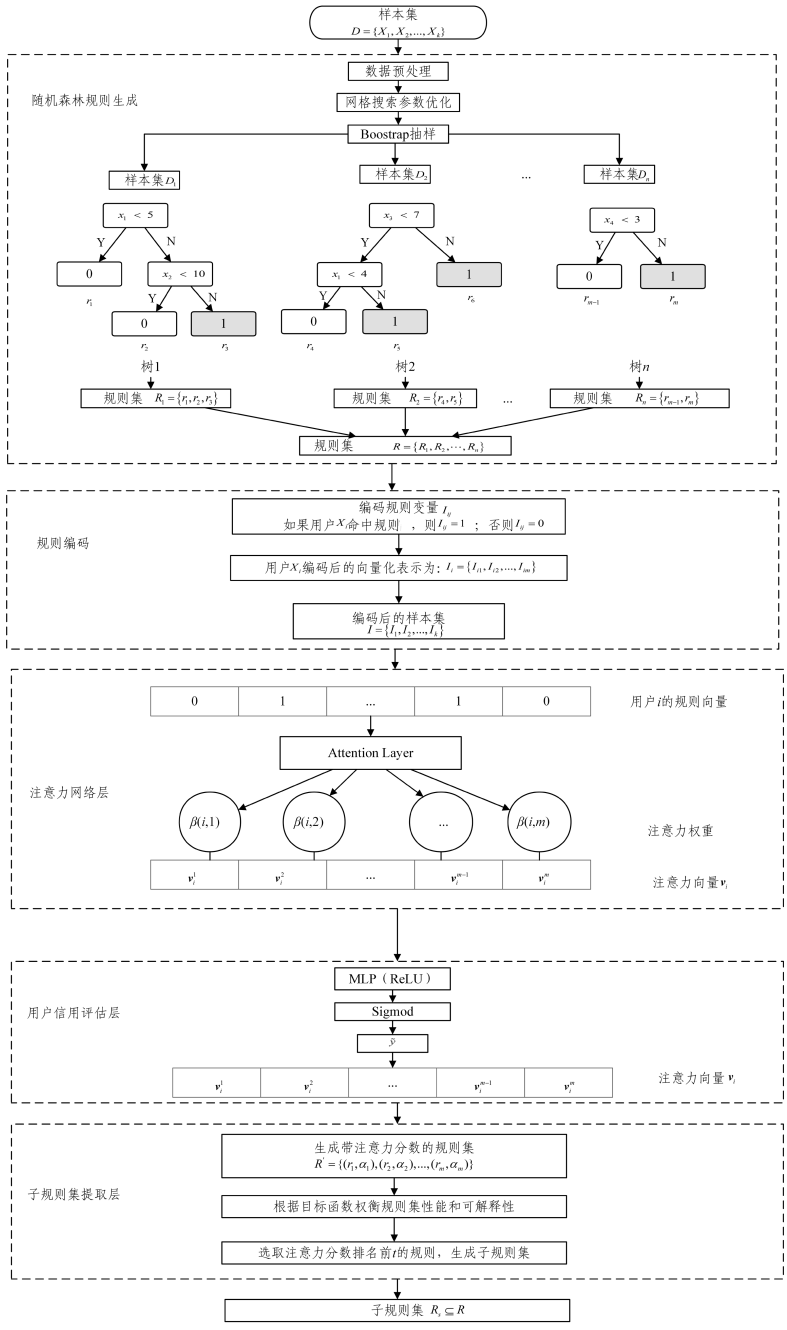

图2 基于注意力机制的可解释用户信用评估模型的框架

Fig. 2 Framework of interpretable user credit evaluation model based on attention mechanism

具体来说，该过程首先从使用样本数据训练随机森林模型开始，以制定一组初始规则。然后对这些规则进行编码，以便进行向量化表示，从而为每个用户的规则生成唯一的向量表示。这些规则的重要性由注意力网络层和用户信用评估层确定。最后，规则子集提取层通过优化一个目标函数来识别最优规则子集，该函数在模型预测准确性和可解释性之间取得平衡。

3.1 随机森林规则提取

在随机森林模型中，每棵决策树都体现了一系列分类或预测规则，这些规则由从根节点到叶节点的路径表示。图 3 给出了一个包含 T 棵决策树的随机森林。根据前述的转换方法，这个随机森林可以转换成规则集 $\boldsymbol{R}$，如表 2 所列。在该规则集中，每条规则对应随机森林中某棵决策树从根到叶的一条路径。规则的条件部分位于左侧，描述了激活该规则必须满足的所有特征约束。相反，标签部分位于右侧，指定了当输入数据满足这些条件时应分配的类别或预测值。随机森林因其复杂且众多的分支而被视为黑箱，这给理解带来了挑战[44]。然而，由它们生成的每个单独分支都是可解释的。本文提出的随机森林规则提取算法如算法 1 所示。

算法 1 集成树规则提取算法 rulesExtract(Trees)

```
输入：Trees // 集成树
输出：R={r1，r2，…，rM} // 规则集
1. R←Ø // 初始化规则集为空
2. Trees // 集成树
3. M← 0 // 初始化规则计数器
4. //遍历每棵树的所有分支
   for each Tree in Trees do
     Branches← extractBranches(Tree) // 提取当前树的所有分支
     for each branch in Branches do
       M← M+1 // 增加规则计数器
       conditions← [] // 初始化决策条件列表
       //遍历分支的每个节点，构建决策条件
       for each node in branch do
           condition← buildCondition(node) // 根据节点构建决策条件
           conditions.append(condition) //将决策条件添加到列表中
           //使用逻辑与连接所有决策条件，形成规则
           rule←"and".join(conditions)
           R.append(rule) //将规则添加到规则集中
     end for
   end for
5. return R //返回规则集
```

由于随机森林中的每棵树都经过独立训练，并且在构建过程中随机选择特征进行分裂，因此所得规则集 R 可能包含大量相似或重复的规则，故规则剪枝变得必要。规则剪枝包括两个部分：1）确定剪枝规则的决策条件，通常采用留一法评估每个规则的必要性，判断其是否应被剪枝[39]；2）优化规则集规模，在确保模型性能的前提下，通过删除冗余规则，最小化规则集的整体规模[41-43]。本文重点探讨规则的最优子集，以提高随机森林的可解释性。鉴于规则决策条件的剪枝不影响本研究，因此视为可选步骤，不予讨论。

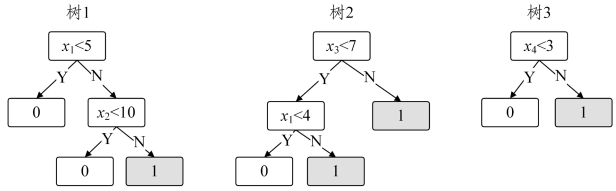

图 3 随机森林模型

Fig. 3 Random forest model

表 2 由随机森林生成的规则集 R

Table 2 Rule set R generated by random forest

规则编号	规则条件	规则结果
r_{11}	$x_1<5$	0
r_{12}	$x_1\geqslant5$ & $x_2<10$	0
r_{13}	$x_1\geqslant5$ & $x_2>10$	1
r_{21}	$x_3<7$ & $x_1<4$	0
r_{22}	$x_3<7$ & $x_1\geqslant4$	1
r_{23}	$x_3\geqslant7$	1
r_{31}	$x_4<3$	0
r_{32}	$x_4\geqslant3$	1

3.2 规则编码

首先，基于生成的规则集构建一个新的数据集 $\boldsymbol{I}$。规则集$\{r_1,r_2,\cdots,r_j\}$的条件集表示为$\{c_1,c_2,\cdots,c_j\}$。$\boldsymbol{I}_{ij}$ 表示条件 c_j 是否满足$\boldsymbol{x}_i$，如式(3)所示：

$$I_{ij}=\begin{cases}1, & \text{如果 } c_j \text{ 满足} \boldsymbol{x}_i\\ 0, & \text{否则}\end{cases} \tag{3}$$

通过将I_{ij}与分类标签相结合，形成了一个新的数据集 $\boldsymbol{I}$：$\{[I_{i1},\cdots,I_{ij},y_i],i=1,\cdots,N\}$。其中，$[I_{i1},\cdots,I_{ij}]$为自变量，$Y$ 为因变量。这种转换将基于规则的分析问题转换为监督学习问题。

3.3 注意力网络层

Liu 等[45]采用自注意力机制方法来提取用户特征，随后运用多层感知机对这些特征进行处理，以预测用户的违约率；Zhao 等[46]通过基于多头注意力机制的 BM-Liner 信用贷款评估模型，解决了词向量固化问题，提高了信贷评估准确率；Zhang 等[47]通过注意力机制实现信息的融合，进而实现文本的分类。受以上文献启发，本文采用注意力神经网络提取重要规则，通过规则编码获得的向量仅由 0 和 1 组成，因此可以将它们输入注意力网络层。用户 k 对应的规则 i 的注意力值通过式(4)计算得出：

$$\alpha(i)=\boldsymbol{h}^{\mathrm{T}}\phi(\boldsymbol{W}_i\cdot\boldsymbol{I}_{ki})+\boldsymbol{b}_1 \tag{4}$$

其中，$\boldsymbol{h}^{\mathrm{T}}$，$\boldsymbol{W}_i$和$\boldsymbol{b}_1$是在过程中训练的参数，$\phi(x)$表示双曲正切(tanh)函数。通过对式(4)进行归一化，得出了在给定规则集中，对于用户 k，规则$\boldsymbol{I}_{ki}$对应的注意力权重值：

$$\beta(k,i)=\frac{\exp(\alpha(i))}{\sum_{i=1}^{m}\exp(\alpha(i))} \tag{5}$$

其中，$\beta(k,i)$表示规则 i 在规则集中的重要性。一旦获得每条规则的注意力权重值 $\beta(k,i)$，用户 k 的注意力向量表示由式(6)定义：

$$\boldsymbol{v}_k=\sum_{i=1}^{m}\beta(k,i)\boldsymbol{I}_{ki} \tag{6}$$

$\boldsymbol{v}_k$作为用户 k 的规则向量表示的加权因子，反映了规则

变量在评估用户信用度时的重要性。具体来说，注意力权重较高的规则变量对$\boldsymbol{v}_k$的贡献更为显著。

3.4 用户信用评估层

用户信用评估层旨在评估用户的信用度。如图 2 所示，用户 k 的规则向量表示为$\boldsymbol{v}_k$，利用多层感知器(MLP)连接层来捕捉用户规则变量之间的相互作用，如式(7)所示：

$$\boldsymbol{r}=MLP(ReLU(\boldsymbol{W}_1 \cdot \boldsymbol{v}_k+b_2)) \tag{7}$$

经过多层感知机后，应用 Sigmoid 函数来评估用户的信用度，如式(8)所示：

$$\tilde{y}=Sigmoid(\boldsymbol{W}_2 \cdot \boldsymbol{r}+b_3) \tag{8}$$

其中，$\boldsymbol{W}_1$，$\boldsymbol{W}_2$，b_2，b_3表示神经网络的参数，$\boldsymbol{r}$ 表示通过全连接层学习到的用户规则向量，$\tilde{y}$表示从用户信用评估层获得的预测值。$\tilde{y}$值为 0 表示用户信用良好，而$\tilde{y}$值为 1 则表明存在潜在的违约风险。

确定用户信用评估的预测值$\tilde{y}$后，本文采用了在二分类任务中经常使用的二元交叉熵(BCE)损失函数，用于量化模型预测的概率分布与实际标签之间的差异，如式(9)所示：

$$\mathcal{L}(y,\tilde{y})=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log(\tilde{y}_i)+(1-y_i)\log(1-\tilde{y}_i)] \tag{9}$$

其中，N 表示样本的总数，y 表示真实的二分类标签值(0 或 1)，$\tilde{y}$表示正类(标签 1)的预测概率，范围在 0～1 间。通过计算式(9)，$\tilde{y}$不断迭代优化逼近真实值，从而优化模型参数。

通过构建上述模型并输入用户信息，可以通过迭代计算确定最优的注意力权重 $\beta(k,i)$。通过对训练集中所有用户的注意力权重求平均，可以得出全局规则注意力权重，并将其作为评估规则集中规则重要性的依据。

3.5 规则子集提取层

规则子集的提取主要关注两个关键维度：规则子集的预测性能及其可解释性。

1) 规则子集的预测性能：通过 AUC、准确率、精确率、召回率和 F1 分数等指标来评估，具体如式(10)所示：

$$\begin{cases} Accuracy=\dfrac{TP+TN}{TP+FP+TN+FN} \\ Precision=\dfrac{TP}{TP+FP} \\ Recall=\dfrac{TP}{TP+FN} \end{cases} \tag{10}$$

其中，TP(真正例)指在评估群体中，用户等级被正确评估的用户数量；TN(真负例)指在评估群体中，用户等级被正确评估为负类的用户数量；FP(假正例)指在评估群体中，用户等级被错误评估的用户数量；FN(假负例)指在所有待评估用户中，被错误评估的用户数量。

AUC 表示 ROC 曲线下的面积，该面积由曲线和坐标轴围成。它用作表征分类器整体性能的指标。AUC 值越大，模型性能越好[48]。准确率表示正确预测的样本占总样本的比例。精确率表示在预测为违约的样本中，实际违约样本的比例。召回率表示在所有实际违约用户中，正确预测为违约用户的比例。

2)规则集的可解释性：在本文中，规则集的可解释性通过规则集中的规则数量来评估，如式(11)所示：

$$Interpretability(\boldsymbol{R}_{\text{sub}})=1-\frac{num\_rule_{\text{sub}}}{num\_rule_{\text{rf}}} \tag{11}$$

其中，num_rule_{sub}表示使用本文方法提取的规则子集中的规则数量，num_rule_{rf}表示由随机森林生成的初始规则集中的规则数量。对于给定的随机森林模型，num_rule_{rf}是恒定的，而num_rule_{sub}越小，表明可解释性越高。

本文中规则集的评估函数如式(12)所示。在用户信用评估领域，AUC 被广泛认可为评估模型预测性能的标准指标[49]。因此，本文采用 AUC 作为评估规则集预测性能的主要指标。$Interpretability_{\text{sub}}$表示规则集的可解释性，$\alpha$ 是[0，1]内的平衡因子。评估函数同时考虑了规则集的预测性能和可解释性。

$$Fitness(\boldsymbol{R}_{\text{sub}})=\alpha\times AUC_{\text{sub}}+(1-\alpha)\times Interpretability_{\text{sub}} \tag{12}$$

规则子集的提取涉及求解式(13)的约束优化问题，这是一个 NP 难问题[26]。本文中规则集中的规则根据其注意力权重按降序排序。然后，基于注意力权重的降序排序直接选取前 num_rule_{sub}条规则构成子集，该策略将 NP 难问题转换为线性时间可解的 O(n)问题(n 为规则数)。

$$\begin{aligned} \max_{R_{\text{sub}}\subseteq R} \ & Fitness(\boldsymbol{R}_{subnum\_rule_{\text{sub}}})=\alpha\times AUC_{\text{sub}}+(1-\alpha)\times \\ & Interpretability_{\text{sub}} \\ \text{s.t.} \ & AUC_{\text{sub}}\geqslant\varepsilon_1 \\ & Accuracy_{\text{sub}}\geqslant\varepsilon_2 \\ & Precision_{\text{sub}}\geqslant\varepsilon_3 \\ & Recall_{\text{sub}}\geqslant\varepsilon_4 \\ & F1_{\text{sub}}\geqslant\varepsilon_5 \\ & Interpretability_{\text{sub}}\geqslant\varepsilon_6 \end{aligned} \tag{13}$$

4 实验过程与结果分析

4.1 数据集

本文使用了用于信用风险评估的 3 个基准数据集进行实验分析。其中 Australian 数据集和 German 数据集来自 UCI，其在信用风险评估算法中被广泛使用[50-51]。LC 数据集来自互联网信贷机构 Lending Club 的真实信贷数据，常被用于信用风险评估算法的实验[52-53]，其样本数超 20 万，正负样本比例严重不均衡，可用于验证本文方法在真实的信贷数据集上的有效性。它们的详细描述如表 3 所列。在模型训练之前，对 3 个数据集进行了欠采样处理，以解决它们存在的类别不平衡问题。

表 3 数据集描述

Table 3 Dataset description

数据集	样本数量	特征数量	非违约数	违约数	数据集来源
Australian	690	14	383	307	http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)
German	1000	20	700	300	http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)
LC	225639	17	177600	48039	https://www.lendingclub.com/info/download-data.action

4.2 实验参数设置

随机森林中的树 *ntree* 为 150 棵，最大深度 *max_depth* 为 6。神经网络采用 Adam 作为模型的优化器，为了降低过拟合的风险并加速收敛，在用户信用评估层部分采用了 Dropout 层和 Batch-Normalization 层来实现对模型的训练优化。2 层信用预测全连接层的维度分别设为 512 和 160。在模型训练时，总的训练 Epoch 为 50，Batch 大小为 92。优化器 Adam 参数设为 0.0002，Dropout 设为 0.3。3 个数据集上的所有实验采用 10-fold 交叉验证。约束条件中的参数默认设置为$\varepsilon_1=0,\varepsilon_2=0,\varepsilon_3=0,\varepsilon_4=0,\varepsilon_5=0,\varepsilon_6=0$。评价函数中的平衡因子 α 设置为 0。

4.3 对比算法的选择

随机森林规则提取算法主要分为基于数学规划的方法和启发式算法[27]。由于数学规划方法的规则提取存在求解效率问题[29-30]，适用于数据集中样本数较少的场景，而在金融信用风险评估的实际场景中，数据集中样本数通常在十万级以上。以本文采用的互联网金融机构 Lending Club 公开的真实用户信贷数据集为例，样本数在 20 万以上，因此在金融信用风险评估的实际应用中，基于数学规划的方法并不适用。

基于启发式算法的规则提取方法只考虑单一的预测性能指标约束（如 AUC，Accuracy）[31-42]，未考虑金融信用风险评估领域数据的特殊性，金融机构的信贷数据中“非违约”样本与“违约”样本比例通常是严重失衡的，“违约”样本数量通常较少，违约样本作为关键少数类，其识别效果直接影响金融机构的风险控制能力。金融机构不仅关注信用风险评估模型的整体准确率，更强调对违约样本的高召回率，即能够有效识别出“违约”样本，否则违约漏检会对金融机构产生重大的资金损失。因此，这类方法在金融信用风险评估场景中不适用。

在信用风险评估实际应用场景中，当前较为广泛使用的 4 种算法分别为：1）RF＋HC，随机森林（RF）结合爬山法（Hill-Climbing，HC），用于提取最优规则子集[38]；2）RF＋HC_CMPR，RF＋HC 的改进版，采用新的规则评估方法[38]；3）RF＋SGL，随机森林结合稀疏群组套索（Sparse Group Lasso，SGL），用于从随机森林转换的规则中提取稀疏权重向量[26]；4）RF＋MSGL，RF＋SGL 的变体，采用多类别稀疏群组套索（Multiclass SGL，MSGL）[26]。这 4 种算法与本文方法都是通过计算规则的重要性来提取规则的，因此本文选用这 4 种算法进行实验对比。

4.4 整体性能对比结果

Australian 数据集上的实验结果如表 4 所列。在对比算法中，AUC 值最高的算法是 RF_HC，其 AUC 值为 0.806，规则数量为 135，因此将式（13）的约束条件中的参数ε_1设置为 0.806，求得的子规则集中规则数量为 63，即本文算法 RRFA 使用了 63 条规则就超过了 RF_HC 算法的预测性能，显著提高了可解释性。

对比算法中，规则数量最少的算法是 SGL，其规则数量是 57，AUC 值是 0.773，将式（13）的约束条件中的参数ε_1设置为 0.773，求得的子规则集中规则数量为 43，即本文算法 RRFA 在达到与 SGL 算法相同的预测性能时，进一步减少了规则的数量，提升了可解释性。

表 4 Australian 数据集上的实验结果

Table 4 Experimental results on Australian dataset

算法	*AUC*	*Accuracy*	*Precision*	*Recall*	*num_rule*
RF	0.871	0.874	0.858	0.850	5 042
RF_HC	0.806	0.786	0.742	0.812	135
RF_HC_CMPR	0.803	0.784	0.747	0.805	123
SGL	0.773	0.760	0.734	0.763	57
MSGL	0.800	0.783	0.758	0.788	74
RRFA(43)	0.778	0.791	0.828	0.669	43
RRFA(63)	0.808	0.815	0.812	0.751	63

German 数据集上的实验结果如表 5 所列。在对比算法中，AUC 值最高的算法是 MSGL，其 AUC 值为 0.624，规则数量为 236，将式（13）的约束条件中的参数ε_1设置为 0.624，求得的子规则集中规则数量为 176，即本文算法 RRFA 使用了 176 条规则就达到了 MSGL 算法的预测性能，显著提高了可解释性。

表 5 German 数据集上的实验结果

Table 5 Experimental results on German dataset

算法	*AUC*	*Accuracy*	*Precision*	*Recall*	*num_rule*
RF	0.653	0.563	0.391	0.873	5667
RF_HC	0.554	0.551	0.362	0.593	146
RF_HC_CMPR	0.582	0.571	0.380	0.630	145
SGL	0.596	0.569	0.372	0.639	94
MSGL	0.624	0.586	0.389	0.652	236
RRFA(70)	0.596	0.567	0.371	0.667	70
RRFA(176)	0.624	0.562	0.380	0.776	176

对比算法中，规则数量最少的算法是 SGL，其规则数量是 94，AUC 值是 0.596，将式（13）的约束条件中的参数ε_1设置为 0.596，求得的子规则集中规则数量为 70，即本文算法 RRFA 在达到与 SGL 算法相同的预测性能时，可以进一步减少规则的数量，提升算法可解释性。

LC 数据集上的实验结果如表 6 所列。在对比算法中，AUC 值最高的算法是 RF_HC_CMPR，其 AUC 值为 0.603，规则数量为 159，将式（13）的约束条件中的参数ε_1设置为 0.603，求得的子规则集中规则数量为 54，即本文算法 RRFA 使用了 54 条规则就达到了 RF_HC_CMPR 算法的预测性能，并且规则数量低于所有的对比算法，显著提高了可解释性。

表 6 LC 数据集上的实验结果

Table 6 Experimental results on LC dataset

算法	*AUC*	*Accuracy*	*Precision*	*Recall*	*num_rule*
RF	0.661	0.576	0.320	0.814	15398
RF_HC	0.599	0.599	0.300	0.591	166
RF_HC_CMPR	0.603	0.603	0.304	0.585	159
SGL	0.602	0.602	0.301	0.611	103
MSGL	0.602	0.602	0.305	0.560	92
RRFA(54)	0.603	0.627	0.305	0.561	54

4.5 实验结果分析

最优子规则集需要综合考虑模型预测性能与可解释性，因此令式（13）中的平衡因子 $\alpha=0.5$。在 3 个数据集上，首先分别按照规则的注意力权重值从高到低对规则进行排序，然后遍历规则集中的规则，计算出子规则集的各项预测性能指标。

Australian 数据集上，随着规则集规模的增长，规则集的

各项性能指标均呈现较为稳定的增长趋势，如图 4 所示。当规则集中规则数量在 100 以内时，各项指标增速较快，表明此时规则集中每增加一条规则，规则集的性能可以得到较高提升。当规则集中规则的数量超过 100 条时，各项指标趋于稳定，表明此时再向规则集中增加规则对规则集的性能提升不显著。此结果表明注意力权重值最高的前 100 条规则，对规则集的预测性能起到决定性作用，剩余的注意力权重值低的规则对规则集的预测性能贡献不显著。

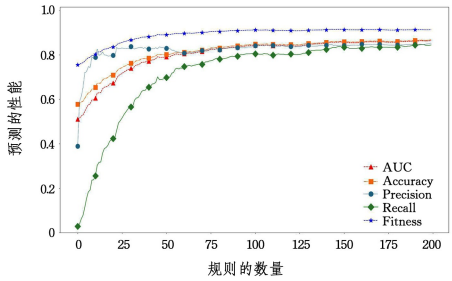

图 4 Australian 数据集子规则集性能指标

Fig. 4 Performance metrics of rule subsets for Australian dataset

German 数据集上，随着规则集规模的增长，规则集中除了 Accuracy 指标，其他各项性能指标均呈现较为稳定的增长趋势，如图 5 所示。当规则集中规则数量在 100 以内时，各项指标变化较为显著，表明注意力权重值最高的前 100 条规则对规则集的预测性能起到决定性作用。当规则集中规则的数量超过 100 条时，各项指标趋于稳定，此时再往规则集中增加规则对规则集的性能提升不显著，表明了注意力权重值低的规则对规则集的预测性能贡献不显著。

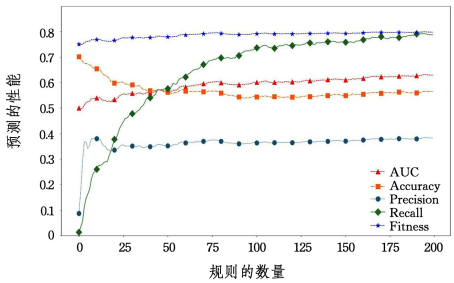

图 5 German 数据集子规则集性能指标

Fig. 5 Performance metrics of rule subsets for German dataset

然而，实验结果表明，当规则数量在 50 以内时，Accuracy 指标随规则数量增加呈现下降趋势，这一现象与本文的直观预期不符。通过深入分析 German 数据集的样本分布特征，发现其正负样本比例存在严重不平衡（违约样本占比仅 30%）。在这种非均衡分布下，若简单地将所有样本预测为多数类（非违约），虽然可以获得表面上的高 Accuracy（理论最大值可达 70%），但会导致少数类（违约）的 Recall 指标归零，这种方法在实际金融信用风险评估场景中是完全失效的。因此，现有的只考虑单一的预测性能指标的规则提取方法在信用风险评估场景中是不适用的[31-42]。本文通过式(13)中的约束条件，可以确保各项指标均满足金融机构的实际需求。

LC 数据集上的结果如图 6 所示，LC 数据集也存在正样本与负样本的比例严重不均衡，其实验结果与 German 数据集上的实验结果类似。

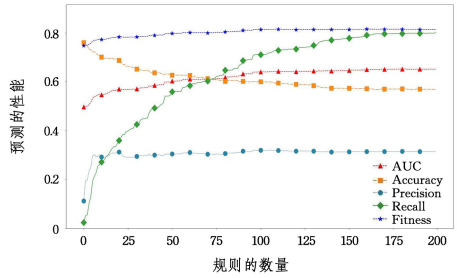

图 6 LC 数据集子规则集性能指标

Fig. 6 Performance metrics of rule subsets for LC dataset

4.6 最小子规则集

当式(13)中的平衡因子 α 为 0 时，可以求得子规则集在满足各项约束条件下的最少规则数量，即满足约束条件的可解释性最高的子规则集。由于规则集中的规则按照注意力权重值由高到低排列，因此遍历规则集中的规则，即可求得最小子规则集。

当式(13)中各约束条件的参数设置为 $\varepsilon_1 = AUC_{rf} \times 0.8$，$\varepsilon_2 = Accuracy_{rf} \times 0.8$，$\varepsilon_3 = Precision_{rf} \times 0.8$，$\varepsilon_4 = Recall_{rf} \times 0.8$，$\varepsilon_5 = F1_{rf} \times 0.8$，$\varepsilon_6 = 0.9$ 时，实验结果如表 7 所列。实验结果表明，在子规则集的各项预测性能指标均达到原始随机森林模型预测性能 80%以上的前提下，在 Australian 数据集上最小子规则集为 46 条规则，在 German 数据集上最小子规则集为 85 条规则，在 LC 数据集上最小子规则集为 86 条规则。

表 7 保持原始 RF 预测性能 80%以上的最小子规则集

Table 7 Minimum rules that maintains at least 80% of the original RF prediction performance

数据集	模型	规则数量	*AUC*	*Accuracy*	*Precision*	*Recall*
Australian	RF	5042	0.871	0.874	0.858	0.850
	RRFA(46)	46	0.792	0.802	0.827	0.701
German	RF	5667	0.653	0.563	0.3911	0.873
	RRFA(85)	85	0.593	0.550	0.363	0.699
LC	RF	15398	0.661	0.576	0.320	0.814
	RRFA(86)	86	0.622	0.604	0.309	0.655

当式(13)中各约束条件的参数设置为 $\varepsilon_1 = AUC_{rf} \times 0.9$，$\varepsilon_2 = Accuracy_{rf} \times 0.9$，$\varepsilon_3 = Precision_{rf} \times 0.9$，$\varepsilon_4 = Recall_{rf} \times 0.9$，$\varepsilon_5 = F1_{rf} \times 0.9$，$\varepsilon_6 = 0.9$ 时，实验结果如表 8 所列。实验结果表明，在子规则集的各项预测性能指标均达到原始随机森林模型预测性能 90%以上的前提下，在 Australian 数据集上最小子规则集为 73 条规则，在 German 数据集上最小子规则集为 188 条规则，在 LC 数据集上最小子规则集为 117 条规则。

表 8 保持原始 RF 预测性能 90%以上的最小子规则集

Table 8 Minimum rules that maintains at least 90% of the original RF prediction performance

数据集	模型	规则数量	*AUC*	*Accuracy*	*Precision*	*Recall*
Australian	RF	5042	0.871	0.874	0.858	0.850
	RRFA(73)	73	0.818	0.825	0.818	0.769
German	RF	5667	0.653	0.563	0.3911	0.873
	RRFA(188)	188	0.625	0.560	0.379	0.786
LC	RF	15398	0.661	0.576	0.320	0.814
	RRFA(117)	117	0.641	0.589	0.316	0.734

在 3 个数据集上，规则集的各项预测性能指标达到随机森林预测性能指定比例时，所需的规则数量如图 7 所示，通过设置式(13)中的约束参数，根据规则注意力权重值由高到低遍历规则集中的规则，即可求得满足约束的最小子规则集。该方法具有重要的现实应用意义。

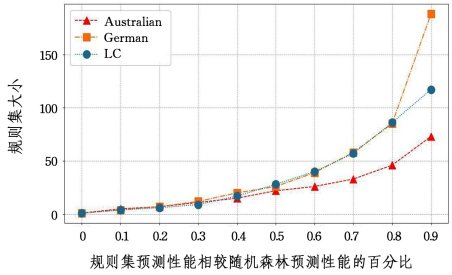

图 7 规则集达到指定预测性能时所需的规则数量

Fig. 7 Number of rules required for the rule set to achieve the specified prediction performance

4.7 计算效率对比

实验硬件环境如下：CPU 为 Intel Core i7-1260P，8 核心，16 线程，主频 2.10GHz，内存为 16GB DDR4。在相同的硬件环境下，本文方法与随机森林及 4 种规则提取算法的计算时间对比如表 9 所列，可以看到，本文方法与 4 种对比算法均具有较高的计算效率，并且在超大规模数据集 LC 上都能够在有限时间内完成求解，具有实际应用意义。本文方法根据注意力权重值将规则降序排序，按照式(13)遍历规则集，可以在 $O(n)$ 时间内得到最优子规则集，其中 n 为子规则集中规则的数量。

表 9 算法计算时间对比

Table 9 Comparison of algorithm computation time

(s)

数据集	*RF_HC*	*RF_HC_CMPR*	*SGL*	*MSGL*	*RRFA*	*RF*
Australian	16	18	14	11	22	1
German	29	36	22	19	38	1
LC	2137	2549	1837	1432	2372	35

结束语 针对信用风险评估中复杂机器学习模型可解释性不足的问题，提出了一种基于注意力权重的规则提取算法，通过优化稀疏性与预测性能的平衡，实现高效透明的信用风险决策。在 3 个公开信用数据集上进行实验，验证了本文方法的有效性，并与随机森林和 4 种主流规则提取方法进行对比。实验表明，本文方法在保证预测性能的同时显著降低了规则复杂度。通过提出的基于注意力权重排序的规则提取框架，可在线性时间内求得最优解，解决了传统方法依赖启发式搜索的高计算成本问题，并通过设计的多性能指标联合约束的优化目标，增强了子规则集的稳定性与业务的适配性。本文方法特别适用于对模型可解释性要求较高的金融领域，例如银行、信贷机构等需要向客户或监管机构提供透明决策依据的应用场景。本文方法通过多指标约束，能够有效处理样本不平衡问题，尤其擅长识别少数类(如违约样本)，从而满足金融机构对高召回率的需求。此外，本文方法采用高效的规则提取与优化策略，能够适应大规模数据集的计算要求，确保其在金融信用风险评估实际落地中的可行性。未来研究将重点提升本文方法的泛化能力，探索其在医疗诊断、工业故障预测等其他高解释性需求领域的应用潜力。同时，还将研究如何结合启发式算法，进一步优化模型的预测性能与可解释性之间的平衡。

参 考 文 献

[1] BAESENS B. Using neural network rule extraction and decision tables for credit-risk evaluation[J]. Management Science, 2003, 49(3): 312-329.

[2] SERRANO-CINCA C, GUTIERREZ-NIETO B. Partial Least Square Discriminant Analysis for bankruptcy prediction[J]. Decision Support Systems, 2013, 54(3): 1245-1255.

[3] MIYAMOTO M, MIYAMOTO M. Credit risk assessment for a small bank by using a multinomial logistic regression model[J]. International Journal of Finance and Accounting, 2014, 3(5): 327-334.

[4] COSTA E SILVA E, LOPES C, CORREIA A, et al. A logistic regression model for consumer default risk[J]. Journal of Applied Statistics, 2020, 47: 1-17.

[5] BEQUÉ A, COUSSEMENT K, GAYLER R, et al. Approaches for credit scorecard calibration: an empirical analysis[J]. Knowledge-Based Systems, 2017, 134: 213-227.

[6] LI T, WANG H, WU J, et al. Sparse Bayesian learning for credit risk evaluation[J]. Journal of Computer Applications, 2013, 33(11): 4.

[7] BHATTACHARYA A, WILSON S P, SOYER R. A Bayesian approach to modeling mortgage default and prepayment[J]. European Journal of Operational Research, 2019, 274(3): 1112-1124.

[8] MELNYK K V, BORYSOVA N V. Improving the quality of credit activity by using scoring model[J]. Radio Electronics Computer Science Control, 2019(2): 60-70.

[9] DAMRONGSAKMETHEE T, NEAGOE V E. Principal Component Analysis and ReliefF Cascaded with Decision Tree for Credit Scoring[M]// Artificial Intelligence Methods in Intelligent Algorithms. Cham: Springer, 2019.

[10] CHERN C C, LEI W U, HUANG K L, et al. A decision tree classifier for credit assessment problems in big data environments[J]. Information Systems and e-Business Management, 2021, 19(1): 363-386.

[11] GOH R, LEE L S. Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches[J]. Advances in Operations Research, 2019, 2019: 1-30.

[12] LEE I G, YOON S W, WON D. A Mixed Integer Linear Programming Support Vector Machine for Cost-Effective Group Feature Selection: Branch-Cut-and-Price Approach[J]. European Journal of Operational Research, 2022, 299(3): 1055-1068.

[13] SHEN F, YANG Z, ZHAO X, et al. Reject inference in credit scoring using a three-way decision and safe semi-supervised support vector machine[J]. Information Sciences, 2022, 606: 614-627.

[14] WANG A Q, HAN Z C, WANG Y L. Risk assessment of logistics finance enterprises based on BP neural network and fuzzy

mathematical model[J]. Journal of Intelligent & Fuzzy Systems, 2020, 39: 5915-5925.

[15] FRAISSE H, LAPORTE M. Return on investment on artificial intelligence: The case of bank capital requirement[J]. Journal of Banking & Finance, 2022, 138: 106401.

[16] KELLNER R, NAGL M, ROSCH D. Opening the black box—Quantile neural networks for loss given default prediction[J]. Journal of Banking & Finance, 2022, 134: 106334.

[17] BHATORE S, MOHAN L, REDDY Y R. Machine learning techniques for credit risk evaluation: a systematic literature review[J]. Journal of Banking and Financial Technology, 2020, 4(1): 111-138.

[18] DASTILE X, CELIK T, POTSANE M. Statistical and machine learning models in credit scoring: A systematic literature survey [J]. Applied Soft Computing, 2020, 91: 106263.

[19] LENKA S R, BISOY S K, PRIYADARSHINI R, et al. Empirical Analysis of Ensemble Learning for Imbalanced Credit Scoring Datasets: A Systematic Review[J]. Wireless Communications and Mobile Computing, 2022, 2022: 6584352.

[20] HOFMAN J M, SHARMA A, WATTS D J. Prediction and explanation in social systems[J]. Science, 2017, 355(6324): 486-488.

[21] CHEN D X, YE J H, YE W C. Interpretable selective learning in credit risk[J]. Research in International Business and Finance, 2023, 65: 101940.

[22] DAVIS R, LO A W, MISHRA S, et al. Explainable Machine Learning Models of Consumer Credit Risk[J]. Journal of Financial Data Science, 2023, 5(4).

[23] DUVNJAK M, MERĆEP A, KOSTANJČAR Z. Intrinsically Interpretable Models for Credit Risk Assessment[C]// 2024 47th MIPRO ICT and Electronics Convention. IEEE, 2024: 31-36.

[24] Equal Credit Opportunity Act[S]. United States Code, title 15, chapter 41, subchapter IV, 1974.

[25] HOOFNAGLE C J, VAN DER SLOOT B, ZUIDERVEEN BORGESIUS F. The European Union general data protection regulation: what it is and what it means[J]. Information & Communications Technology Law, 2019, 28(1): 65-98.

[26] MASHAYEKHI M, GRAS R. Rule extraction from decision trees ensembles: new algorithms based on heuristic search and sparse group lasso methods[J]. International Journal of Information Technology & Decision Making, 2017, 16(6): 1707-1727.

[27] HADDOUCHI M, BERRADO A. A survey and taxonomy of methods interpreting random forest models [J]. arXiv: 2407.12759, 2024.

[28] MARTENS D, BAESENS B, GESTEL T V, et al. Comprehensible credit scoring models using rule extraction from support vector machines[J]. European Journal of Operational Research, 2007, 183(3): 1466-1476.

[29] HADDOUCHI M, BERRADO A. Forest-ORE: Mining an optimal rule ensemble to interpret random forest models[J]. Engineering Applications of Artificial Intelligence, 2025, 143: 109997.

[30] BIRBIL S I, EDALI M, YUCEOGLU B. Rule Covering for Interpretation and Boosting[J]. Information Fusion, 2020, 63: 196-207.

[31] MANZALI Y, ELFAR M. Optimizing the number of branches in a decision forest using association rule metrics[J]. Knowledge and Information Systems, 2024, 66(6): 3261-3281.

[32] BORUAH A N, BISWAS S K, BANDYOPADHYAY S. Transparent rule generator random forest (TRG-RF): an interpretable random forest[J]. Evolving Systems, 2023, 14(1): 69-83.

[33] BOLOGNA G. A rule extraction technique applied to ensembles of neural networks, random forests, and gradient-boosted trees [J]. Algorithms, 2021, 14(12): 339.

[34] EDALI M. Performance analysis of set partitioning formulations on the rule extraction from random forests[J]. Pamukkale University Journal of Engineering Sciences, 2021, 27(4): 513-519.

[35] CHEN M, HUO J, DUAN Y. An interpretable model for sepsis prediction using multi-objective rule extraction[J]. Journal of Intelligent Information Systems, 2024, 62(5): 1403-1429.

[36] SHAMS Z, DIMANOV B, KOLA S, et al. REM: An Integrative Rule Extraction Methodology for Explainable Data Analysis in Healthcare[R]. medRxiv, 2021.

[37] WANG S, WANG Y, WANG D, et al. An improved random forest-based rule extraction method for breast cancer diagnosis[J]. Applied Soft Computing Journal, 2020, 86: 105941.

[38] MASHAYEKHI M, GRAS R. Rule extraction from random forest: the RF+HC methods[M]// Advances in Artificial Intelligence. Cham: Springer, 2015: 223-237.

[39] DENG H. Interpreting tree ensembles with intrees[J]. International Journal of Data Science and Analytics, 2019, 7(4): 277-287.

[40] DONG L, YE X, YANG G. Two-stage rule extraction method based on tree ensemble model for interpretable loan evaluation [J]. Information Sciences, 2021, 573: 46-64.

[41] FRIEDMAN J H, POPESCU B E. Predictive learning via rule ensembles[J]. The Annals of Applied Statistics, 2008, 2(3): 916-954.

[42] DUMITRESCU E, SULLIVAN H, HURLIN C, et al. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds[J]. Working Papers, 2021.

[43] KATO H, HANADA H, TAKEUCHI I. Safe rulefit: Learning optimal sparse rule model by meta safe screening[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2330-2343.

[44] LEI X N, LIN L F, XIAO B Q, et al. Re-exploration of small and micro enterprises' default characteristics based on machine learning models with SHAP[J]. China Journal of Management Science, 2024, 32(5): 1-12.

[45] LIU X Y, QU Y W, ZHOU Q Y. Self-attention credit evaluation model [J]. Computer Engineering and Applications, 2019, 55(13): 36-41.

[46] ZHAO X F, WU D L, WU W W, et al. BM-Linear credit loan evaluation model based on multi-head attention mechanism[J]. Journal of Systems & Management, 2023, 32(1): 118.

[47] ZHANG M Q, ZHOU H, CAO J G. Directed sentiment text

classification based on attention mechanism and dual BERT[J]. CAAI Transactions on Intelligent Systems，2022，17(6)：1220-1227.

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

[49] VERBRAKEN T，BRAVO C，WEBER R，et al. Development and application of consumer credit scoring models using profit-based classification measures[J]. European Journal of Operational Research，2014，238：505-513.

[50] QIAN X，CAI H H，INNAB N，et al. A novel deep learning approach to enhance creditworthiness evaluation and ethical lending practices in the economy [J]. Annals of Operations Research，2025，346：1597-1619.

[51] YANG F，ABEDIN M Z，HAJEK P. An explainable federated learning and blockchain-based secure credit modeling method [J]. European Journal of Operational Research，2024，317(2)：449-467.

[52] XIA Y，JIANG S，MENG L，et al. XGBoost-B-GHM：An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring [J]. Systems，2024，12(7)：254.

[53] TRINH L T. A co mparative analysis of consumer credit risk models in Peer-to-Peer Lending [J]. Journal of Economics，Finance and Administrative Science，2024，29(58)：346-365

WANG Baocai，born in 1988，postgraduate. His main research interests include machine learning interpretability and intelligent credit risk control systems.

WU Guowei，born in 1973，Ph.D，professor，Ph.D supervisor，new century outstanding talents of Ministry of Education，Executive member of CCF System Software Special Committee. His main research interests include advanced computing and intelligent systems.

（责任编辑：喻藜）