基于自动化特征组合的隐私保护风险识别机制

蔡民超[1,2]，姚宏伟[1]，王旸[1]，秦湛[1]，陈少梦[2]，任奎[1]
（1.浙江大学网络空间安全学院，浙江 杭州 310007；2.杭州快迪科技有限公司，浙江 杭州 310000）

摘　要：异常行为识别（AD）算法在实际应用中，通常会面临特征组合优化困难、分类器准确率难提高、模型应用效率低等技术挑战。用户所产生的多维数据具有丰富的空间结构信息，围绕这些多维数据的特点，在通过同态加密的隐私保护方式进行数据脱敏的基础上，针对特征组合优化困难的技术挑战，提出并实现了首个基于特征分箱的自动化特征组合优化模型算法，该算法在特征组合优化方面提升了99.93%的计算效率。基于自动化特征组合优化模型筛选出的重要特征所组合的规则仍存在分类器准确率难提高的技术挑战，故将自动化筛选出的重要特征融入识别模型中，设计并实现了首个规则和算法的交叉应用模型，并将该方式应用到基于用户多维信息的异常行为识别中，在识别先享不付类异常用户的具体场景中实现资金挽损效率提升27.78%。

关键词：异常行为识别；多维信息；风险识别；可信度模型；同态加密
中图分类号：TP181
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024194

Privacy protection risk identification mechanism based on automated feature combination

CAI Minchao[1,2], YAO Hongwei[1], WANG Yang[1], QIN Zhan[1], CHEN Shaomeng[2], REN Kui[1]
1. School of Cyber Science and Technology, Zhejiang University, Hangzhou 310007, China
2. Hangzhou Kuaidi Technology Co., Ltd., Hangzhou 310000, China

Abstract: In practice, the anomaly detection (AD) algorithm usually faced technical challenges such as difficulty in optimizing feature combinations, difficulty in improving classifier accuracy, and low model application efficiency. The multi-dimensional data generated by users was with rich spatial structure information, revolved around the characteristics of the multidimensional data. Building upon the privacy protection method using homomorphic encryption, the technical challenge of optimizing feature combinations was addressed. The first automated feature combination optimization model algorithm based on feature binning was proposed and implemented. This algorithm enhanced computational efficiency in feature combination optimization by 99.93%. The rules combined by the important features selected by the automatic feature combination optimization model still faced the technical challenge of difficulty in improving the classifier accuracy. Therefore, the important features selected automatically were integrated into the recognition model, the first cross-application model of rules and algorithms was designed and implemented. This approach was applied to anomaly detection based on multi-dimensional user data, resulting in a 27.78% increase in funds saved in the specific scenario of identifying abnormal users who enjoy first but do not pay.

Keywords: anomaly detection, multi-dimensional information, risk identification, trustworthiness model, homomorphic encryption

收稿日期：2024–05–31；**修回日期：**2024–09–30
通信作者：秦湛，qinzhan@zju.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2021YFB3100300）；国家自然科学基金资助项目（No.U20A20178, No.62072395, No.62206207）
Foundation Items: The National Key Research and Development Program of China (No.2021YFB3100300), The National Natural Science Foundation of China (No.U20A20178, No.62072395, No.62206207)

0 引言

随着大数据的蓬勃发展，行为识别已成为机器学习、人机交互、模式识别、数据挖掘等领域的研究热点[1-2]。异常行为识别（AD, anomaly detection）作为当今信息科技中数据处理的重要技术[3]，是一种旨在检测和预防可能发生异常行为或事件的识别机制。该机制在多个领域中具有广泛的应用前景，随着数据规模的不断增大和机器学习、深度学习等技术的不断发展，异常行为识别机制在网络安全、金融风控、工业生产等领域被广泛部署并深入研究。

异常行为识别也随着信息技术的发展和需求而逐渐趋于复杂，通过特征组合形成单一分类器的方式已经无法满足异常行为的检测需求。例如，用户在基于地理位置的服务平台进行交互时，会产生具有丰富空间结构信息的多维数据实体，因而基于地理位置服务平台的异常行为识别系统通常会采用Hadoop、Spark等平台对流量大数据进行预处理，并引入统计分析方法或基于机器学习的异常识别模型[4-5]，但以上方式仍存在特征选用方式单一、特征组合优化困难等问题。

许多研究会采用机器学习的异常行为识别模型从原始特征中提取有效信息，然而在高维、样本多样、数据结构复杂的数据集中，该类模型无法获得理想的数据潜在分布[6]。尤其在样本数据严重不平衡或缺少异常标记的数据集上，模型检测性能不高[7]，这使得单纯的机器学习模型较难满足具有复杂空间结构特性数据集的异常检测需求。

针对以上的特征组合优化困难问题，本文研究了特征寻优的过滤方法、包装方法、嵌入方法等特征选择算法技术，并针对具有复杂空间结构的订单特征集进行数据实验，设计并实现了一种基于特征分箱的自动化特征筛选方法，该方法可较好地通过剪枝降低计算量，并实现特征选用的价值量化。针对分类器准确率难提升、识别机制应用效率低等问题，本文研究了规则组合与机器学习在异常检测中的应用技术，并针对基于地理位置服务应用平台的先享不付场景进行实验，设计并实现了一套规则和算法的交叉应用模型，该模型有效提升了分类器准确率难提升的问题，并通过AB实验实现数据结构复杂的应用场景中的价值量化，在识别先享不付类异常用户的具体场景中实现资金挽损效率提升27.78%。在解决识别机制应用效率低的问题上取得了显著结果。

本文的主要贡献包括以下3个方面。

1) 提出了一种基于特征分箱的自动化特征筛选方法，能够通过剪枝降低计算量，并实现良好的特征价值量化。

2) 设计了一种规则和算法的交叉应用模型，将自动化筛选出的重要特征融入新的交叉应用模型中，提高了分类器的准确率。

3) 对某个基于地理位置的服务应用展开实验验证，实验结果表明本文设计的方法能够有效地提高识别效率和准确率，并在实际应用中带来了具体价值提升。

1 预备知识和相关研究工作

1.1 异常

异常是一种由各种与常规不符的异常行为所产生的现象或事件，随着技术的发展，不同的作者对异常有不同的定义。Kaur等[8]认为异常为离群值被认为是一组定义在簇外的噪声点，也可以定义为位置外同时也与噪声分离的点。Doostari等[9]认为社交网络中的异常应定义为偏离大多数观察的观察结果。结合已有定义和对网络异常产生方式的分析，本文中的异常是指在特定网络中个人或群体的行为不符合正常模式定义的特征行为。

1.2 异常行为识别

异常行为识别机制是安全风险中进行风险感知识别和风险决策的核心工具，是一种用于自动化决策和执行的实时计算体系。它具有自动化决策、实时响应、复杂业务高速处理等优势，所以该识别机制被广泛应用于互联网、金融等多种安全风险检测领域。异常行为识别机制基于预定义的规则和算法能够快速分析和处理大量数据流，并根据规则和策略进行决策。这些规则和策略可以根据数据特性和应用需求进行定义和配置，从而实现高效的异常检测和决策管理。

异常行为识别机制的功能会因为应用场景和数据体系的不同而不同，通常的策略体系需要以规则匹配为主，可定制串行化和并行化策略，并且会依据应用需求增加AB实验模块、熔断降级模块、侦测日志模块和命中统计模块等。

异常行为识别机制基于识别方式的不同通常可

以分为统计类、规则类、监督学习类、机器学习类等。作为安全风险识别机制的重要方法，国内外许多学者对异常行为检测进行了研究。Liu 等[10]提出了一个对比自监督学习框架（CoLA, contrastive learning to localize actions），该框架使用图神经网络和对比学习方法，充分利用网络中的本地信息，提高了属性网络中异常检测的性能，并能够灵活地适应大型网络。Chen 等[11]提出了一种新颖的异构 BiGAN-based 异常检测模型 HTA-GAN，该模型结合一类分类器和创新的异常评分函数，通过生成对抗网络（GAN, generative adversarial networks）结构来有效捕捉时间特征并计算实际的异常分数，从而提高了无监督异常检测的性能和鲁棒性。Wu 等[12]提出了一种新型异常检测模型，通过二维离散小波变换和自注意力机制有效提高了模型在图像级异常检测中的性能。Li 等[13]总结归纳了深度学习在多变量时间序列异常检测中的应用，分类讨论了异常时间点、时间区间和时间序列的检测技术，并指出了当前方法在异常解释性方面的不足。以上研究极大地丰富了异常行为识别的方法和应用场景，但仍存在识别结果可解释性差、应用效率难提升等挑战。

为了处理复杂空间结构的数据内容，研究者们将机器学习和数据挖掘相关技术引入网络异常行为检测中，利用机器学习方法对异常行为进行聚类判断。在传统机器学习领域，机器学习模型方法包括 LightGBM[14]、XGBoost[15]、K-means 等，该类方法一般将时序问题转换为监督学习问题，通过特征工程和机器学习算法进行异常行为检测。机器学习的数据挖掘统计模型主要分为生产式和判别式 2 种[16-18]，前者通过学习联合概率，侧重分析各类数据的分布情况，后者则研究条件概率，关注各类数据分类的边界。在进行异常流量或行为检测时，除了采用有监督类和无监督类机器学习方法外，也可采用基于支持向量机的聚类算法[19-20]进行异常流量或行为判断。但上述方法对样本的要求都较为严格，在面对具有复杂空间结构的数据集建模时，通常会存在适配性差、准确率难提升等问题。

面对上述研究中所存在的问题，本文提出了一种基于特征分箱的自动化特征筛选方法和规则与算法交叉应用模型，有效解决了识别结果可解释性差、准确率难提升的技术挑战，并将其应用于识别基于地理位置服务应用平台的异常用户中，取得了资金挽损效率提升 27.78% 的显著成果，有效突破了模型应用效率难提升的技术挑战。

2　识别机制设计

针对特征组合优化困难、分类器准确率难提高、模型应用效率低三大技术挑战，本文设计了一套异常行为识别机制，包含识别流程设计、策略体系设计和评价体系设计三部分。该机制首先针对行为发生过程中的实时数据进行预处理，通过实时计算方式实现特征加工和特征组合，并将组合后的策略应用于异常行为识别和决策中。该安全风险识别机制具备实时计算、特征加工、实时识别等功能，可以有效解决特征组合优化困难、计算量过大等问题，并利用基于该机制的策略评价体系和数据应用效果评估来实现系统分类器准确率的优化，可有效解决当前存在的准确率难提升和应用效率低的相关问题。本文使用的缩略词表如表 1 所示。

表 1　　本文使用的缩略词

缩略词	定义
TP	true positive 的缩写，真实行为样本且未触发异常行为策略的用户请求次数
TN	true negative 的缩写，真实行为样本且触发异常行为策略的用户请求次数
FN	false negative 的缩写，异常行为样本且触发异常行为策略的用户请求次数
FP	false positive 的缩写，异常行为样本但未触发异常行为策略的用户请求次数
ROC	receiver operating characteristic 的缩写，即受试者工作特征曲线
AUC	area under the curve 的缩写，即曲线下的面积
KS	kolmogorov-smirnov 的缩写，是统计学中的一种非参数假设检验
WOE	weight of evidence 的缩写，是一种区分正负样本能力的统计指标
IV	information value 的缩写，是一种衡量特征预测能力的统计指标

2.1　识别流程设计

本文基于用户多维数据具有复杂空间结构信息的特性，设计了一种安全的风险识别方法，该方法主要分为行为发生时的实时计算和行为发生后的离线分析两部分。在行为发生时，该方法利用行为埋点数据和第三方数据引入每个事件桩点进行特征加

工和组合，并将组合后的特征与模型平台中的模型结果进行交叉合并，进而形成策略中心中的策略，实现对每个事件桩点中行为识别的决策和干预。在行为发生后，该方法将数据中心存储的历史行为埋点数据应用于离线模型和规则的应用和评价中。异常行为识别流程如图1所示。

在异常行为识别机制中，通过对行为埋点数据等数据源的加工和预处理，以及模型平台中模型输出值的再加工，可以产出特定场景所需的特征。通过特征的组合和筛选来实现异常行为识别系统中异常行为的检测和决策，并通过评价体系和专家经验来提升异常检测和决策的能力和应用价值。

以上识别机制针对用户多维数据具有丰富空间结构信息的特性，在现有的国内外异常行为识别方法的基础上，采用实时数据流和离线数据流同步处理同步加工的方式，可有效提升数据应用效率，并采用评价体系和特征组合筛选方法来降低数据计算复杂度，在兼顾特征组合效率的同时也有效降低了计算资源的消耗。该识别机制也创新性地采用模型输出值作为特征之一的方式将模型和规则实现了交叉应用，将模型输出值应用于特征组合和筛选中，较好地提升了识别机制的灵活性，同时可以有效突破单一规则准确率难提升的问题。

以上识别机制运行方式的形成与基于地理位置服务应用平台的发展有一定关联。面对地理位置、操作路径、设备信息、交易记录等各类数据形成的复杂空间结构信息，传统的单一维度分析方法难以全面捕捉和理解这些行为模式。该识别机制采用多维数据处理方法，能够更精细地分析和识别异常行为，提升识别的准确性。在复杂的应用场景中，风险事件通常需要在极短的时间内进行识别和决策，故该识别机制通过实时计算和决策，能够在行为发生时立即进行分析和干预，确保风险控制的及时性和有效性。该识别机制通过策略中心实现策略的集中管理和动态调整，结合实时数据和离线数据的同步处理，能够根据最新的数据分析结果不断优化和更新识别策略，提升了识别机制的适应性和持续改进能力。而应用场景中数据量的增加和复杂度的提升对计算资源提出了更高要求，该识别机制采用评价体系和特征组合筛选的方法降低数据计算复杂度，在提升特征组合效率的同时，显著降低了计算资源的消耗。

2.2　策略体系设计

基于安全风险中具有丰富空间结构信息的多维数据，在此将数据预处理后加工成相应特征，并通过特征组合的方式形成策略组，进而通过策略组的分类方式实现异常行为的识别和检测。行业内通常会通过专家经验从海量特征中实现特征的筛选和组合，但这种方式缺少科学的寻优方法和量化逻辑。故本文提出了一种基于特征分箱的自动化特征筛选方法，该方法可较好地通过剪枝来降低计算量，并实现特征选用的价值量化。需要从事件相关的上百

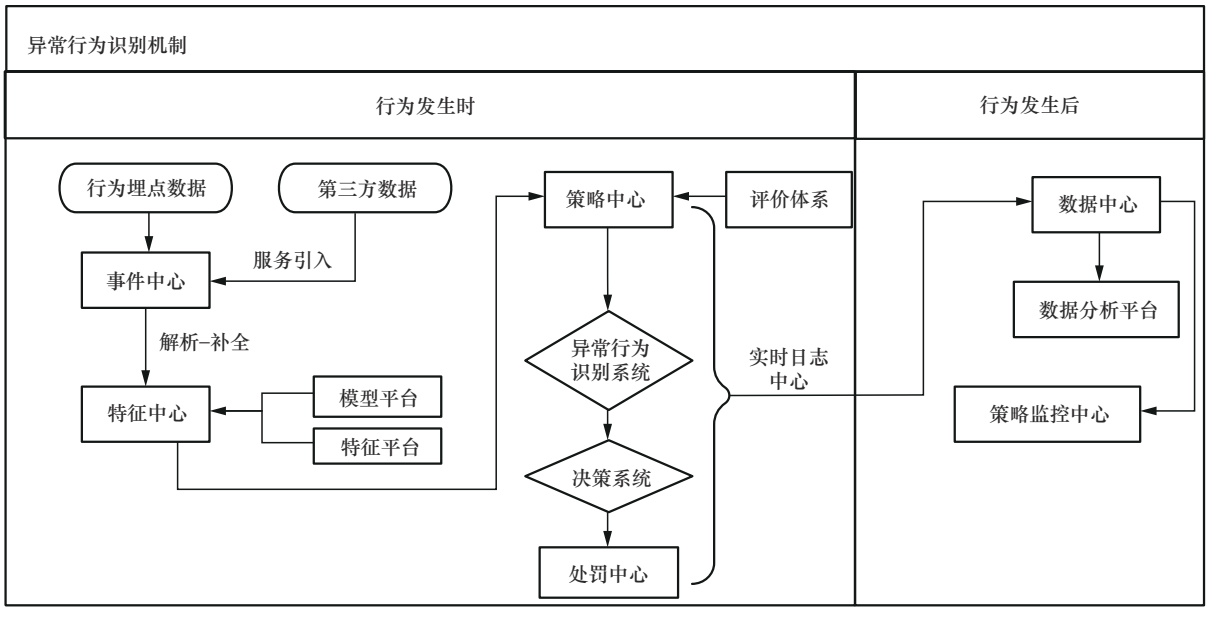


图1　异常行为识别流程

项特征中筛选目标特征并进行特征组合，解决特征寻优时长和特征组合效果量化问题。

在实际模型建设时，本文发现仅通过特征组合方式形成的策略组在基于用户多维数据的复杂场景中，极易出现分类器准确率难以提升的问题，但仅使用机器学习模型的方式又极易出现识别结果可解释性差的问题。基于以上问题，在多个弱分类器组合成强分类器的过程中，需要探寻一种可保证分类器准确率且对强分类器具有较强可解释性的方法。因此，本文设计了一种基于特征分箱的自动化特征组合优化算法和一种规则和算法的交叉应用模型和评价框架，首先将单一特征或多特征经过特征分箱的自动化特征组合优化算法实现目标特征的筛选，并将组合后的特征应用于有监督算法模型中，将算法模型应用于基于特征组合的策略组中，该方式在具有较高的识别结果可解释性的同时可以较好地解决分类器准确率难提升等问题。具体的策略体系设计架构如图2所示。

在基于特征分箱的自动化特征组合优化算法中，对于连续特征X，本文将其分割成k个区间，每个区间表示为B_i，其中i=1,2,⋯,k。分箱的目的是将连续变量转换为分类变量，以便更好地进行后续分析。

在规则和算法的交叉应用模型中，假设输出的全局最优价值$V=\max\{v_0,v_{y1},v_{y2},\cdots,v_{yn}\}$，其中全局最优价值的衡量单位为V，原规则组合后对全局输出的价值为v_0，模型的数据结果值为y，模型应用至各个核心规则后全局输出的价值为v_{yn}，则y应用至各个核心规则后全局最优的价值$V=\max\{v_0,v_{y1},v_{y2},\cdots,v_{yn}\}$。通过如上的规则与模型的交叉应用方式可以较好地确定出全局最优的交叉方案。

2.3　评价体系设计

在特征组合选用和异常识别结果评估方面，均需要一套体系化的评价方法来实现基于用户多维信息的风险识别数据安全机制的应用量化。因本文中的识别机制本质上是一个二分类问题，主要通过多策略分类器来实现异常行为检测。为此，本文提出了一套针对识别结果评估和特征选用及组合的评价体系，以实现对识别机制效果和价值的综合评价。

2.3.1　策略评价指标

异常行为识别是一个经典的二元分类问题，通常需要使用混淆矩阵表示分类结果，用矩阵的列表示类的实例预测，矩阵的行表示类的实例。一个典型的混淆矩阵如表2所示。

表2　　混淆矩阵

实际值	预测为真	预测为假
实际为真	TP（真阳性）	TN（真阴性）
实际为假	FP（假阳性）	FN（假阴性）

准确率（Accuracy）是最常用的分类性能指标，可以用来表示模型的精度，即模型识别正确的个数/样本的总个数。一般情况下，模型的精度越高，说明模型的性能越好。

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FN}+\text{FP}} \tag{1}$$

但在异常行为这种类别分布明显不平衡的场景中，仅用Accuracy难以真实评价模型性能的好坏，且无法区分出模型在FN（假阴性）与FP（假阳性）的判错差异。因此需要考虑使用敏感性（Sensitivity）和特异性（Specificity）来进一步辅助评价模型性能的好坏。

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP}+\text{TN}} \tag{2}$$

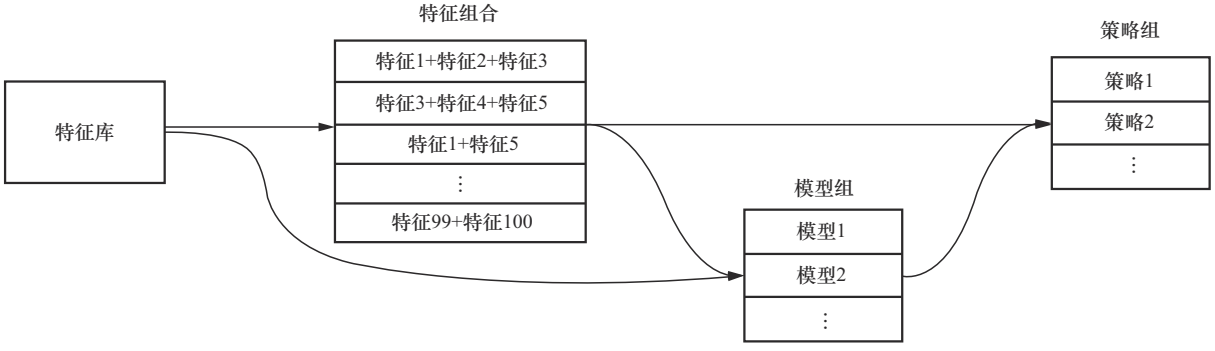


图2　策略体系设计架构

敏感性真阳性率（TPR, true positive rate），该值越高，判定为TN（真阴性）的占比越小，即被误判定为异常行为的真实行为比例越低。

$$\text{Specificity} = 1 - \frac{\text{FP}}{\text{FP} + \text{FN}} \tag{3}$$

特异性（Specificity），即1−假阳性率（FPR, false positive rate），该值越高，判定为FP（假阳性）的占比越小，即被误判定为真实行为的异常行为比例越低。

然而采用敏感性与特异性作为验证策略的评测指标时，仅能反映分类结果的好坏而无法对分类过程进行调优。故在此引入受试者工作特征（ROC, receiver operating characteristic）曲线与曲线下的面积（AUC, area under the curve）进行策略评价。ROC曲线描述的是分类混淆矩阵中FPR与TPR这2个变量之间的变化趋势。

在策略选用过程中，为了更好地区分正负样本，实现高精准低误伤的识别能力，故在此引入指标KS（kolmogorov-smirnov），KS是用于衡量正负样本比例差程度大小的评估指标。通过好坏样本之间累计分布的差值评估策略的风险区分能力。好坏样本的累计差异越大，模型评估策略的风险区分能力越强，KS指标越大。KS值的取值范围是[0,1]，KS越大，表示计算预测值的模型区分好坏用户的能力越强。KS的表达式如式(4)所示。

$$\text{KS} = \left| \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{TN} + \text{FP}} \right| = |\text{TPR} - \text{FPR}| \tag{4}$$

2.3.2 特征评价指标

在进行策略配置或模型训练之前需要对特征进行筛选过滤，为提升策略应用效果应尽可能选用对预测结果有重要影响的变量而剔除影响不大的变量，并防止模型出现过拟合的现象。在此引入区分正负样本能力的指标证据权重（WOE, weight of evidence）和衡量特征预测能力的指标信息价值（IV, information value）。

WOE是一种对特征分箱进行编码的方式，能够赋予每一个分箱一个编码值，该值不仅能够标识一个分箱，还能够代表分箱对于预测的贡献，表达式如式(5)所示。

$$\text{WOE}_i = \ln\left(\frac{b_i}{b}\right) - \ln\left(\frac{g_i}{g}\right), i = 1,2,\cdots,10 \tag{5}$$

其中，b_i和g_i分别表示分箱中坏样本和好样本的数量，b和g分别表示总体样本中坏样本和好样本的总数量。WOE是一种反映某个变量预测能力的指标，从计算的角度来看，WOE可以计算出特征每一个分箱的预测水平，但WOE无法计算特征整体的预测能力或者对于预测的贡献。故在此引入可以计算特征整体预测能力的指标IV，该指标有效弥补了WOE中整体得分是各个分箱得分简单求和的缺陷。将WOE的加权求和定义为IV值，计算式如式(6)和式(7)所示。

$$\text{IV}_i = \left(\frac{b_i}{b} - \frac{g_i}{g}\right)\text{WOE}_i, i = 1,2,\cdots,10 \tag{6}$$

$$\text{IV}_i = \left(\frac{b_i}{b} - \frac{g_i}{g}\right)\left(\ln\left(\frac{b_i}{b}\right) - \ln\left(\frac{g_i}{g}\right)\right), i = 1,2,\cdots,10 \tag{7}$$

整个特征的IV值则为每段IV值之和，计算式如式(8)所示。

$$\text{IV} = \sum_{i}^{n}\text{IV}_i, i = 1,2,\cdots,10 \tag{8}$$

特征的IV值越大，表明该特征的信息价值就越大，对于判断用户好坏的贡献越大，一般也认为这样的特征更适合入模。

传统的特征选用方式非常依赖于统计或专家经验，缺少科学的评价方法和价值量化。本文将特征筛选和策略评价进行融合，引入基于混淆矩阵、ROC曲线和AUC值、KS值为评价指标的策略评价体系，并引入基于WOE和IV值为评价指标的特征评价体系。在策略评价方面，传统的策略评价体系可能仅依赖于单一指标，而通过引入ROC曲线、KS值等，形成了多维度的评价体系。这样的多维度评价不仅能够提升策略评价的科学性，还能从多角度确保策略效果的最大化，提升风险识别和决策的准确性。在特征选用方面，本文创新性地引入指标WOE和指标IV实现特征预测能力的量化，WOE能够细化到每个分箱的预测贡献，IV则能够综合评估整个特征的预测能力。通过这2个指标的结合，可以更精准地筛选出对模型预测结果有显著影响的特征，从而提高模型的预测精度和可靠性。

引入ROC曲线、KS值、WOE、IV值等评价指标，将策略评价体系与特征选用评价体系进行融合，不仅形成了全面科学的评价框架，还提升了特征筛选和模型性能的准确性和有效性。这种多维度的评价方法，结合专家经验和数据驱动，实现了特

征选用和策略应用效果的最大化，创新性地提高了异常行为识别的精度和灵活性。

3　安全风险识别机制应用

3.1　建模准则

本文基于国内某地理位置服务相关应用平台的多维用户信息，将上文提出的风险识别数据安全机制进行应用。在基于地理位置服务的互联网平台上，为了让服务者和被服务者获取更优质的交易体验，通常会向被服务者提供“先享后付”的服务并为服务者提供“完单即结算”的垫付服务。但很多被服务者在被服务结束后并未及时还款导致平台出现应收款缺口，进而形成坏账。为了提前识别该类被服务者，平台通常会依据被服务者和服务者的交易特征和历史行为特征进行特征组合来实现虚假交易的事前、事中和事后识别。如何从特征组合中获取最优特征组合以及如何对特征组合结果进行量化成为获取最优识别策略的核心挑战。

本次应用中的策略组主要由特征组合和模型两部分组成。在特征组合上，主要基于用户的实时或离线的行为数据和平台埋点数据来设计特征项，通过用户特征组合来实现策略组建设，进而实现满足分类器准确率条件的异常数据请求的精准识别。在模型上，主要是将用户的行为信息、订单信息、平台留痕信息等数据通过数据标准化处理、特征组合与衍生等方法构建用户特征，并通过数据分箱、特征选择等特征工程方法，应用到逻辑回归、KNN算法、随机森林等二分类算法模型中，进而将模型应用于安全风险的识别中。

3.2　实验环境配置

本文实验环境基于macOS操作系统，使用Python语言编写，Python版本号为3.11.0，在某O2O公司自研的机器学习平台进行模型搭建，实验环境CPU采用Intel Core i5 2GHz 4核，内存大小为16 GB。

3.3　数据准备

在数据准备阶段，本文获取了实验所需的对应数据，并确保了数据的完整性、准确性和一致性。对数据通过清洗和预处理、数据隐私保护、特征加工等过程来实现数据在建模过程的可用性。

3.3.1　实验数据描述

以国内基于地理位置服务的某应用平台作为实验主体，获取某个时间段中547 692笔已完成服务的订单，其中306 166笔订单未发生逾期，131 197笔订单逾期天数为1至29日，790笔订单逾期天数为30日或30日以上。在此平台中，若被服务者订单逾期天数大于或等于30日，通常会认定为该笔订单为“坏账订单”，即平台将会承担该笔订单导致的资金损失。本文将通过建模对被服务者的未来订单是否会发生逾期进行预测，若被服务者行为被模型归类为异常行为，则被服务者再次发单时系统将会强制要求被服务者进行预付。

本文应用该平台中服务者和被服务者的所有行为数据集和订单特征数据集，将547 692笔订单所涉及的被服务者特征、服务者特征和订单特征中的数百项特征进行特征分箱，并通过基于特征分箱的自动化特征组合优化模型进行特征筛选。

在筛选出可应用于可信模型建设的重要特征后，将通过规则与算法的交叉应用模型来实现异常订单的识别模型建设。在验证模型价值的对照实验中，将流量划分为命中规则与算法交叉模型且命中后强制预付、命中规则与算法的交叉模型但命中后未强制预付和未命中规则与算法的交叉模型三类，并进行实验观察。

3.3.2　数据隐私保护

由于该应用平台中的实验数据含有大量用户隐私信息，而传统的异常检测方法往往需要访问原始数据，因此存在隐私泄露的风险。本文引入了同态加密算法中的Paillier加密对部分用户和平台的敏感数据进行加密，如订单金额、用户手机号等。Paillier加密是一种基于整数群的公钥加密方案，它支持加法同态性质，可以在加密状态下进行加法操作而不需要解密。Paillier加密具有加法同态性质，即2个密文的乘积对应着2个明文和的加密结果。

在密文状态下，本文所提基于特征分箱的自动化特征组合优化模型和规则与算法的交叉应用模型可以在不暴露原始数据的情况下发挥算法作用，并识别出潜在的异常行为。在异常检测完成后，可以使用解密密钥将结果解密，得到原始的异常检测结果。

针对基于多维用户信息的复杂数据内容，使用同态加密可以在保证数据隐私的情况下进行异常识别分析。尤其针对本文应用场景中实时数据流识别机制，利用实时数据流处理技术对加密数据进行实时分析和处理，可以在保护信息安全的

同时及时检测和响应异常行为。通过结合同态加密等隐私保护方法，以及基于规则、机器学习、实时数据流处理等异常行为识别方法，可以构建一个高效且安全的多维数据融合分析系统。这种融合不仅能够保障用户数据的隐私，还能提高异常行为识别的准确性和实时性，为各类应用场景提供了强有力的安全保障。

3.4 基于特征分箱的自动化特征组合优化模型

基于用户多维数据的识别机制因具有复杂空间结构信息的特性，所以进行特征选择是实现异常识别的重要步骤，Al-tashi等[21]对特征选择相关问题和挑战进行系统性文献综述，并批判性地分析了用于解决该问题的建议技术。特征选择是一项重要的数据预处理过程，针对基于安全风险中具有丰富空间结构信息的多维信息，通常会采用数据预处理后加工成相应特征，并通过过滤式特征选择方式形成策略组，进而通过策略组作为分类器来实现异常行为的识别和检测。特征选择是一个NP-hard组合优化问题，假设一个n维数据集，可能存在$2n$的特征子集，遍历所有特征子集的组合方案显然是不可能的。在历史的研究中存在大量搜索方法来解决特征选择的子集组合问题，如顺序向前选择（SFS, sequential forward selection）、顺序向后选择（SBS, sequential backward selection）[22]等。然而这些方法仍然未解决计算量大和解释性不佳的问题。

从事件相关的上百项特征中筛选目标特征并进行特征组合，存在解决特征寻优时长问题、特征组合效果量化问题等相应挑战，面对复杂的特征组合问题，在此创新性地提出了一种基于特征分箱的自动化特征组合寻优方法。该方法在高质量特征推荐、降低计算成本等方面具有显著效果。通过特征分箱、特征多维度处理大幅降低计算难度，并通过WOE、IV值、准确率、召回率等方式实现特征与标签的关联性量化，有效提升策略的筛选效率和筛选准确度。

3.4.1 样本提取与特征预处理

在样本提取期间，将订单id作为主键，依据订单id来获取该订单中服务者和被服务者的各类行为特征和订单特征，并对订单是否未支付超过30日进行打标，将需要识别的目标样本打标为1，否则打标为0。

针对预选的所有特征进行特征预处理，依据缺失值是否过多、数值或枚举值是否单一对低质量特征进行删除，这些低质量特征对策略低贡献或无贡献，删除后可降低计算资源的消耗。

3.4.2 特征分箱与组合

特征分箱是数据预处理中重要的一环，其主要目的是将连续型特征离散化，以便于后续的策略应用和模型训练。通过将连续取值的特征进行分箱处理，可以简化模型的计算复杂度，同时提高特征的区分度和解释性。

将原始的单个特征进行分箱处理，每个箱作为策略所需的最小粒度的条件单元，样本在该最小特征单元上的值只有1或0这2种枚举值，如基于地理位置服务平台中订单的订单距离D，常规取值范围为[0,+∞)，现将该取值范围依据业务规则经验和订单分布表现分成4个箱，分别为[0,190)、[190,500)、[500,1 000)和[1 000,+∞)。每个箱有且仅代表一个特征，特征值只有1或0这2种枚举值，1代表落在该区间内，0代表落在该区间外。本文在实际实验验证中，考虑到各特征的特性，会采用多种方法确定分箱规则，如使用等距分箱、等频分箱或基于业务规则和经验的分箱等方法。

为了体现原始特征不同取值对标签的识别能力的影响，将特征值进行拆分合并，这里用WOE对各区间的标签识别能力进行量化。WOE方法是信用评分卡常用的分箱方案，本文创新性地将该方法应用于基于地理位置服务的应用平台中，因该类平台含有大量基于地理位置的原始特征，如订单距离、订单速度、订单加速度等，该方法可以有效区分该类特征取值的标签识别能力。

以订单距离为例，将这一原始特征样本转化成分箱后的特征样本，其中“标签tag”表示该订单是否为未付超30日订单，转化内容如表3和表4所示。

表3 原始特征样本

订单id	标签tag	订单距离
1	0	100
2	1	450
3	0	20 000
4	0	200
5	0	700

表4　　分箱后的特征样本

订单id	标签tag	订单距离			
		[0, 190)	[190, 500)	[500, 1 000)	[1 000, +∞)
1	0	1	0	0	0
2	1	0	1	0	0
3	0	0	0	0	1
4	0	0	1	0	0
5	0	0	0	1	0

完成分箱转化后需要对特征进行相应组合，进而实现将低准确率的单特征通过组合得到高准确率的策略。例如，表5和表6展示了从2个特征组合到3个特征组合的案例。

表5　　2个特征组合案例

订单id	标签tag	bin_*a*&bin_*b*	bin_*x*&bin_*y*
1	0	1	0
2	1	0	0
3	0	0	0
4	0	1	0
5	0	0	1

表6　　3个特征组合案例

订单id	标签tag	bin_*a*&bin_*b*&bin_*c*	bin_*x*&bin_*y*&bin_*z*
1	0	0	0
2	1	0	0
3	0	0	0
4	0	1	0
5	0	0	0

以上随机组合的复杂度将会随着参与组合特征数的增加而增加，以200维单特征为例，如果从2个特征组合到5个特征组合进行遍历计算，则每一次新组合的计算量都是一次指数级别的增加，最大计算量为$C_{200}^{2}+C_{200}^{3}+C_{200}^{4}+C_{200}^{5}$=2 601 668 290次。

为降低计算量和特征选用复杂度，本文发现大部分特征组合互相之间相关性较高且部分特征组合的标签识别能力很弱，因此可以在进行下一轮组合之前，对当前的特征组合进行一次相关性计算和准确率、召回率、IV值的计算，并剔除准确率低、召回率小、IV值小等特性的低信息价值特征。针对高相关的2个特征，本文将剔除信息价值低的特征。假设在从2个特征组合逐渐增加到5个特征组合的过程中，每次都通过上述方式保留排名前200的特征，则总的计算量将会变为$C_{200}^{2}+200\times198+200\times197+200\times196$=138 100次。

计算量通过特征选用评价的方式从原来的26亿余次降低至138 100次。但在实际的应用场景中，不可能将138 100项特征组合结果发布至系统中进行逐一评估，故本文针对计算结果采用以下方式进行创新性量化。

1) 针对全部的组合后规则，计算每个特征出现的次数，找出次数前3的特征，如(f_1,f_2,f_3)，若(f_1,f_2,f_3)中任何一个特征出现次数小于2，则终止。

2) 从组合后的规则中筛选出包含(f_1,f_2,f_3)的所有规则。

3) 针对已产出规则，根据准确率、召回率和IV值进行评分排序。

本文可根据排序后的规则进行先后尝试，与以上创新方式相比采用特征价值评价方式处理后的筛选方式，计算量从138 100次降至数十几次以内，计算效率提升99.93%，计算量和计算复杂度大幅降低。以上方法在基于地理位置服务的应用平台中均有较好效果，例如，在虚假交易导致的未付场景中，在现有策略的基础上，以30日未付为标签，进一步产出预付策略。头部的单条策略准确率远高于策略组的平均准确率，且规则特征有较好的可解释性。

3.4.3　可信模型建设

为将异常行为检测中的用户可信度和交易行为可信度进行量化，本文采用逻辑回归模型。逻辑回归是一种可以用来分类的常用统计分析方法，并且可以得到概率型的预测结果，属于一种概率型非线性回归[23]。逻辑回归算法作为数据挖掘、数据分类应用中最常用的方法之一[24]，在语义网服务匹配[25]、流行病研究[26]、预测地质灾害[27-28]等方面有着广泛的应用。Ohlson[29]首次利用Logistic回归搭建信用分类模型，分类效果明显。Dinh等[30]的研究结果表明，逻辑回归模型是传统信用风险研究中预测精度较高的模型，具备很强的稳健性。

本实验为提升模型应用的对比效果，也尝试引入GBDT（gradient boosting decision tree）模型等方法进行价值提升，但在处理基于用户多维信息的海量特征时，GBDT模型在树的数量多和树的深度

较大时更容易发生过拟合，需要复杂的参数调节来缓解过拟合问题。并且在具体应用场景中进行异常行为识别时，需要所应用模型具有较好的可解释性，而GBDT模型相比逻辑回归难以解释每个特征的具体贡献和模型的整体决策过程。

本文主要使用Python编程语言来实现整个实验流程，包括数据预处理、特征工程、模型训练、模型评估和结果展示等，使用pandas库进行数据加载、清洗和转换，使用WOE和IV进行特征选择与编码，使用scikit-learn库中的逻辑回归类进行模型定义和训练。

本文采用典型二分类模型−逻辑回归模型作为主要的实验方法，针对异常行为识别进行了详细的算法实现。具体实现步骤包含数据清洗、特征选择与特征工程、数据标准化、模型训练、模型评估等。本文在引入逻辑回归方法的基础上，创新性地引入基于特征分箱的自动化特征筛选方法来实现特征提取，并引入多维特征评价体系实现重要特征选用和策略优化，并在模型应用过程中使用规则与算法的交叉应用模型实现应用价值最优。

本文采用逻辑回归的方式对基于地理位置服务相关平台中服务者和被服务者的可信度进行量化，其中被服务者可信度如式(9)和式(10)所示。

$$P_c = h_\theta(x_c) = g(\theta_c^t x_c) = \frac{1}{1+\mathrm{e}^{-\theta_c^t x_c}} \tag{9}$$

$$\theta_c^t x_c = \theta_{1c}x_{1c} + \theta_{2c}x_{2c} + \cdots + \theta_{nc}x_{nc} \tag{10}$$

其中，x_{nc}和θ_{nc}分别为被服务者c的第n个特征和第n个特征的权重，特征内容包含被服务者的行为数据、订单数据和与服务者的交互行为数据，如单位时间成单次数、与同一服务者交互频次等。为构建逻辑回归模型，需要对训练集数据进行类标记。若用户在近期某时间段内P_c低于设定阈值则认定为被服务者a的类标签为1，否则为0。服务者可信度的量化过程与被服务者可信度量化过程相似，采用P_d表示，x_{nd}特征内容包含被服务者的轨迹数据、订单数据和被服务者的交互行为数据，如多单轨迹异常率、与同一被服务者交互频次等。

在判断订单是否为虚假交易时，也会采用类似的逻辑回归算法进行订单真实性的评估。主要的引用特征包含当前订单服务时长和距离、当前订单与服务者或被服务者历史订单时长和距离的差异、是否是定向交易、服务者与被服务者是否有历史关联性、服务者集中性交易比例等。除此之外，也会引入上文中的P_c和P_d作为量化被服务者和服务者的可信度值一同作为订单是否为虚假交易的重要特征。例如，订单m利用以上方法重新构建模型并测试，该模型可表示为

$$P_m = h_\theta(x_m) = g(\theta_m^t x_m) = \frac{1}{1+\mathrm{e}^{-\theta_m^t x_m}} \tag{11}$$

$$\theta_m^t x_m = \theta_{1m}x_{1m} + \theta_{2m}x_{2m} + \cdots + \theta_{nm}x_{nm} \tag{12}$$

3.4.4 识别机制应用

以国内基于地理位置服务的某应用平台作为实验主体，获取某个时间段中547 692笔已完成服务的订单，其中306 166笔订单未发生逾期，131 197笔订单逾期天数为1至29日，790笔订单逾期天数为30日或30日以上。在此平台中，若被服务者订单逾期天数大于或等于30日，通常会认定该笔订单为“坏账订单”，即平台将会承担该笔订单导致的资金损失。基于以上背景，得到实验数据集如表7所示。

表7 实验数据集

逾期天数	订单数
0	382 708
（0,30）	163 997
[30,+∞)	987

因以上场景将采用二分类模型进行应用，故逾期天数区间为（0,30）的样本将作为灰样本，训练与测试期间未使用。在此以训练集80%，测试集20%进行分层拆分，其中实际训练时按照0/1比例为10进行欠采样。训练集和测试集标签分布如表8所示。

表8 训练集和测试集标签分布

标签	训练集样本量	测试集样本量
0	306 166	76 542
1	790	197
−1	131 197	32 800

在特征筛选环节，基于地理位置服务的平台具有多维空间结构数据的特性，所以地理位置信息、服务者历史行为、被服务者历史行为、历史订单数据和当前行为数据均可作为特征筛选的基础数据。在选用特征方面采用上文中基于特征分箱的自动化特征组合优化方法，从系统中筛选出经过计算初筛

后的特征。因数据采集平台的信息安全合规要求，被服务者相关特征在此不做列举，在此仅列举与平台服务者相关的12个变量如表9所示。

表9　　服务者重要特征和权重系数

特征	权重系数
服务者前7日完成订单的实收金额	0.314
服务者年龄	0.111
服务者历史订单数量	0.035
服务者被被服务者7日投诉单数	0.012
服务者端连续不在线天数	0.070
服务者平台分数	0.081
服务者两单时间间隔秒(本次与上次)	0.114
服务者最近7日应用平台服务平均间隔	−0.024
服务者最近7日订单平均金额	0.059
服务者最近7日订单平均支付间隔	0.559
服务者最近30日订单数	0.019
服务者最近7日与90日订单总金额比例	−0.123

在算法训练过程中，采用逻辑回归方式5折交叉验证调参，训练时训练集和验证集AUC分布结果对比如图3所示，训练集和验证集KS分布结果分别如图4和图5所示。

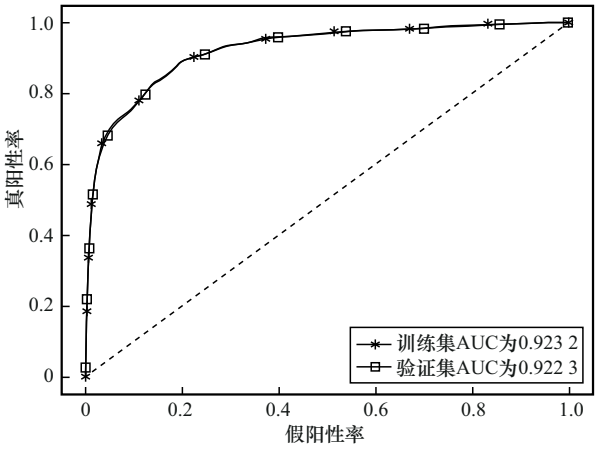


图3　训练集和验证集AUC分布结果

完成模型训练后，利用测试集进行模型测试，为验证模型在订单是否逾期的二分类问题中的区分能力，故将标签0与−1均归类到标签0，即只要发生订单逾期则归为标签0。测试集中AUC分布和KS分布结果如图6和图7所示。

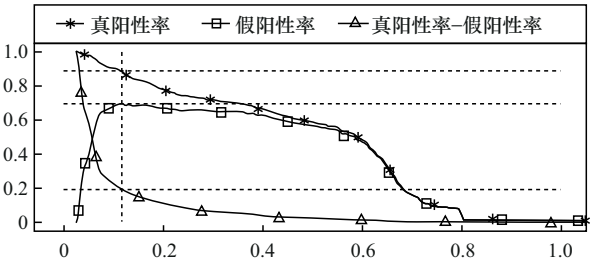


图4　训练集KS分布结果(训练集KS为0.694 4,阈值为0.116)

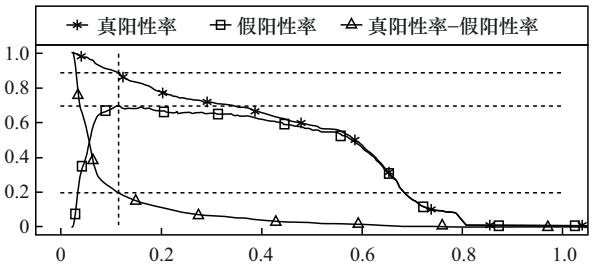


图5　验证集KS分布结果(验证集KS为0.693 8,阈值为0.115)

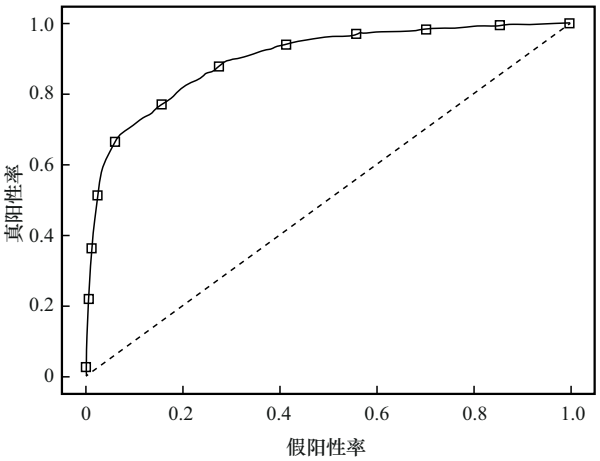


图6　测试集AUC分布结果(验证集AUC为0.898 0)

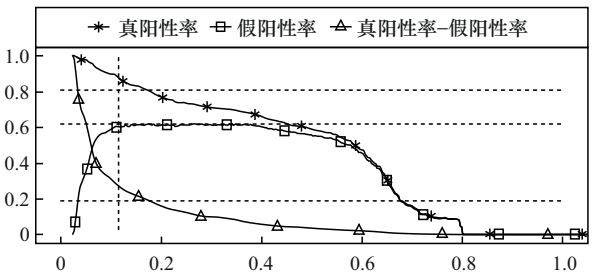


图7　测试集KS分布结果(验证集KS为0.618 3,阈值为0.176)

模型在不同阈值下全量订单覆盖占比与未付订单覆盖占比的表现如图8所示，结合模型评分趋势与业务专家经验，当总订单覆盖率等于0.8%、未付订单覆盖率等于18%时，模型表现效果最佳。故在此选用3个不同但是接近的阈值来作为模型部

署时的灰度阈值，选用50%流量作为模型应用的实验流量，剩余的50%流量作为对照组。具体打分阈值和灰度比例如下。

1) 打分阈值1：0.655，灰度比例为10%。

2) 打分阈值2：0.660，灰度比例为20%。

3) 打分阈值3：0.670，灰度比例为20%。

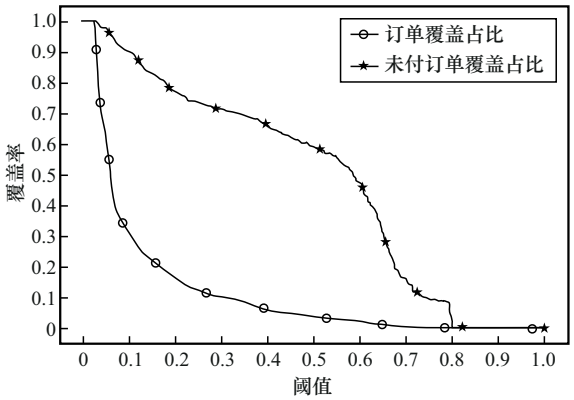


图8 不同阈值下订单覆盖占比与未付订单覆盖占比分布

50%的自然流量应用于文中的实验组，实验组采用规则和算法的交叉应用模型对实验流量进行分类。50%的自然流量应用于对照组，对照组采用传统纯规则方式对实验流量进行分类。基于如上的阈值设定和AB实验分组，整理模型部署后8日的实验组和测试组数据，其策略命中率表现数据表如表10所示。

阈值1的灰度比例为10%，预估在全量范围内策略命中率为0.8%。阈值2的灰度比例为20%，预估在全量范围内策略命中率为0.69%。阈值3的灰度比例为20%，预估在全量范围内策略命中率为0.615%。在上文中，当策略命中率为0.8%时，模型表现效果最佳，故在此选用阈值1作为本模型的最终全量模型上线。

原有的基于特征组合形成的策略组中所有规则已在线上进行实时识别和决策，所有规则之间采用“或”的串行关系，一旦命中规则之一即定义该行为为异常行为，同一行为会请求所有规则执行判断，针对异常行为导致的未付订单会在被服务者发起订单前进行识别并强制用户进行预付，预付金额通常低于订单的实际总金额。

为验证此模型在原有策略组的增益情况和对平台坏账金额产生的影响，故需要观察阈值1在灰度总流量10%后的未付订单表现，当被服务者发起订单时，该部分流量中订单一旦满足阈值1标准则需要强制用户预付，下面以模型上线后订单结束16日仍未付的订单数据为例进行价值测算，结果如表11所示。

在此将未付率指标，即未付金额与应收金额的比值，作为衡量实验组与对照组对于模型应用价值的标准。在测算模型全量放开情况下的异常行为识别收益时，为直接体现模型上线前后的价值变化，特进行如下对照试验。

将经过策略组识别的流量划分为命中规则与算法的交叉应用模型且命中后强制预付、命中规则与算法的交叉应用模型但命中后未强制预付和未命中规则与算法的交叉应用模型三类。为减少实验对线上数据的影响，命中规则与算法的交叉应用模型且命中后强制预付的流量仅生效10%，命中规则与算法的交叉应用模型但命中后未强制预付的流量仅生

表10 实验组和测试组策略命中率表现数据

时间	总体请求量	阈值1（灰度比例为10%）		阈值2（灰度比例为20%）		阈值3（灰度比例为20%）	
		阈值1命中量	阈值1命中率	阈值2命中量	阈值2命中率	阈值3命中量	阈值3命中率
第1日	343 522	337	0.098%	574	0.167%	526	0.153%
第2日	333 869	389	0.117%	487	0.146%	509	0.152%
第3日	304 950	240	0.079%	356	0.117%	294	0.096%
第4日	308 488	232	0.075%	368	0.119%	371	0.120%
第5日	317 489	200	0.063%	399	0.126%	330	0.104%
第6日	320 775	210	0.065%	447	0.139%	386	0.120%
第7日	331 385	264	0.080%	513	0.155%	380	0.115%
第8日	334 098	212	0.063%	446	0.133%	402	0.120%
总计	2 594 576	2 084	0.080%	3 590	0.138%	3 198	0.123%

表 11　　未付订单价值表现数据

序号	分组	单量	总应收金额	未付应收金额	未付应收中预付金额	未付单量	未付率
1	策略命中且灰度预付(开放 10%)	139	3 820.50	98.85	62.00	2	0.96%
2	策略命中但未灰度预付(放过,对照,50%)	751	21 663.91	1 618.34	0.00	37	7.47%
3	其他	168 701	3 901 210.65	7 344.68	293.00	262	0.18%

效 50%，并进行实验观察。基于此计算阈值 1 扩大至全量实验数据后对总体订单中未付率提升价值和对照组扩大至全量实验后对总体订单中未付率提升价值。经试验计算，在对照组中，仅采用纯规则方式进行流量分类且未应用规则和算法的交叉应用模型的场景中未付率为 0.260 8%，在实验组中，使用规则与算法交叉应用模型的场景中未付率为 0.188 36%，即新交叉应用模型上线后未付率下降 27.78%。

该实验以国内基于地理位置服务的某应用平台作为实验主体，面对具有复杂空间结构特性的数据集，在进行特征筛选时，采用基于特征分箱的自动化特征组合优化模型将计算量从 138 100 次降至数十次以内，计算效率提升 99.93%。通过逻辑回归模型实现服务者和被服务者的可信度模型建模，并通过规则与算法的交叉方法应用于平台未付用户的预测和预付决策中，通过对照实验证明，该模型的应用有效促使该场景未付率下降 27.78%。

4　结束语

在以上实验过程中，本文基于特征分箱的自动化特征组合优化方法来解决特征繁多而优化困难的问题，并以此快速定位目标特征。通过二分类相关模型来实现异常行为识别的模型建设，并通过混淆矩阵等评价体系实现对模型准确率等性能的全面提升。在模型应用层面，本文针对行业内普遍存在的模型应用效率低问题，将基于自动化特征组合优化方法筛选出重要特征融入识别模型中，并将规则与算法进行交叉应用，创新性地通过 AB 实验及对比方式来确定全局最优的模型阈值，进而实现策略组在实际业务价值中的效率提升和价值实现。

该方式相比于传统基于专家规则的实时风控体系具有计算成本低、识别准确率高、应用效率高等优势。在计算成本方面，该风险识别机制通过特征组合优化实现 99% 以上的计算效率提升。在识别准确率方面，该风险识别机制通过规则算法的交叉应用模型和体系化评价机制有效提升了模型准确率。在应用效率方面，本文通过将风险识别机制针对性地应用于具有丰富空间结构的数据体系中，实现应用效率提升 27.78%。

本文主要针对实际场景数据进行研究，以解决基于地理位置服务应用中实际业务问题为驱动，进行应用创新和价值创造。本文在引入数据集验证方面和引入多模型对比实验方面仍存在一定不足，将在未来工作中引入更多数据集和更多模型进行实验验证，以进一步验证方法的有效性和鲁棒性。此外，本文还计划改进现有算法，以提升其在不同场景中的适用性和性能。同时，将探索更多的特征工程和模型优化技术，以解决当前算法在实际应用中遇到的瓶颈和挑战。最终目标是构建一个更全面、更高效的异常行为识别系统，能够广泛应用于各种复杂的实际业务场景中。

参考文献：

[1] PINZ A, ZISSERMAN A, WILDES R P, et al. What have we learned from deep representations for action recognition? [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7844-7853.

[2] TRAN D, WANG H, TORRESANI L, et al. A closer look at spatiotemporal convolutions for action recognition[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6450-6459.

[3] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[J]. ACM Computing Surveys, 2009, 41(3): 1-58.

[4] AHMED M, MAHMOOD A N, HU J K. A survey of network anomaly detection techniques[J]. Journal of Network and Computer Applications, 2016, 60: 19-31.

[5] ZHAO Y, NASRULLAH Z, LI Z. PyOD: a python toolbox for scalable outlier detection[J]. arXiv Preprint, arXiv: 1901.01588, 2019.

[6] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]//Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2008: 413-422.

[7] ZIMEK A, SCHUBERT E, KRIEGEL H P. A survey on unsupervised outlier detection in high-dimensional numerical data[J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2012, 5(5): 363-387.

[8] KAUR R, SINGH S. A survey of data mining and social network analysis based anomaly detection techniques[J]. Egyptian Informatics Journal, 2016, 17(2): 199-216.

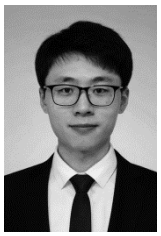
[9] DOOSTARI M, ZEINALI R, LASHKARI H, et al. Anomaly detection in cliques of online social networks using fuzzy node-fuzzy graph[J]. Journal of Basic and Applied Scientific Research, 2013, 3(8):614-626.

[10] LIU Y X, LI Z, PAN S R, et al. Anomaly detection on attributed networks via contrastive self-supervised learning[J]. IEEE Transactions

on Neural Networks and Learning Systems, 2022, 33(6): 2378-2392.
[11] CHEN P, LIU H Y, XIN R Y, et al. Effectively detecting operational anomalies in large-scale IoT data infrastructures by using a GAN-based predictive model[J]. The Computer Journal, 2022, 65(11): 2909-2925.
[12] WU K, ZHU L, SHI W H, et al. Self-attention memory-augmented wavelet-CNN for anomaly detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(3): 1374-1385.
[13] LI G, JUNG J J. Deep learning for anomaly detection in multivariate time series: approaches, applications, and challenges[J]. Information Fusion, 2023, 91: 93-102.
[14] KE G L, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems, 2017, 30: 3149-3157.
[15] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 785-794.
[16] 杜德慧, 程贝, 刘静. 面向安全攸关系统中小概率事件的统计模型检测[J]. 软件学报, 2015, 26(2): 305-320.
DU D H, CHENG B, LIU J. Statistical model checking for rare-event in safety-critical system[J]. Journal of Software, 2015, 26(2): 305-320.
[17] 姚潍, 王娟, 张胜利. 基于决策树与朴素贝叶斯分类的入侵检测模型[J]. 计算机应用, 2015, 35(10): 2883-2885.
YAO W, WANG J, ZHANG S L. Intrusion detection model based on decision tree and naive-Bayes classification[J]. Journal of Computer Applications, 2015, 35(10): 2883-2885.
[18] 马江洪, 张文修, 徐宗本. 数据挖掘与数据库知识发现: 统计学的观点[J]. 工程数学学报, 2002, 19(1): 1-13.
MA J H, ZHANG W X, XU Z B. Data mining and knowledge discovery in database: a statistical viewpoint[J]. Chinese Journal of Engineering Mathematics, 2002, 19(1): 1-13.
[19] WAHAB O A, MOURAD A, OTROK H, et al. CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks[J]. Expert Systems with Applications, 2016, 50: 40-54.
[20] BIGDELI E, MOHAMMADI M, RAAHEMI B, et al. A fast and noise resilient cluster-based anomaly detection[J]. Pattern Analysis and Applications, 2017, 20(1): 183-199.
[21] AL-TASHI Q, ABDULKADIR S J, RAIS H M, et al. Approaches to multi-objective feature selection: a systematic literature review[J]. IEEE Access, 2020, 8: 125076-125096.
[22] LIU H, YU L. Toward integrating feature selection algorithms for classification and clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(4): 491-502.
[23] MUSA A B. Comparative study on classification performance between support vector machine and logistic regression[J]. International Journal of Machine Learning and Cybernetics, 2013, 4(1): 13-24.
[24] MAALOUF M. Logistic regression in data analysis: an overview[J]. International Journal of Data Analysis Techniques and Strategies, 2011, 3(3): 281-299.
[25] 许冲, 戴福初, 徐素宁, 等. 基于逻辑回归模型的汶川地震滑坡危险性评价与检验[J]. 水文地质工程地质, 2013, 40(3): 98-104.
XU C, DAI F C, XU S N, et al. Application of logistic regression model on the Wenchuan earthquake triggered landslide hazard mapping and its validation[J]. Hydrogeology & Engineering Geology, 2013, 40(3): 98-104.
[26] WEI D P, WANG T, WANG J. A logistic regression model for semantic web service matchmaking[J]. Science China Information Sciences, 2012, 55(7): 1715-1720.
[27] ZHANG Z, LIU A, LYLES R H, et al. Logistic regression analysis of biomarker data subject to pooling and dichotomization[J]. Statistics in Medicine, 2012, 31(22): 2473-2484.
[28] JUNEK W N, JONES L W, WOODS M T. Use of logistic regression for forecasting short-term volcanic activity[J]. Algorithms, 2012, 5(3): 330-363.
[29] OHLSON J A. Financial ratios and the probabilistic prediction of bankruptcy[J]. Journal of Accounting Research, 1980, 18(1): 109-131.
[30] DINH T H T, KLEIMEIER S. A credit scoring model for Vietnam's retail banking market[J]. International Review of Financial Analysis, 2007, 16(5): 471-495.

[作者简介]

蔡民超（1991– ），男，山东烟台人，浙江大学博士生，主要研究方向为数据安全、隐私保护、反欺诈等。



姚宏伟（1993– ），男，福建泉州人，浙江大学博士生，主要研究方向为可信人工智能、大模型安全等。

王旸（1988– ），男，浙江嘉兴人，浙江大学博士生，主要研究方向为网络安全、数据安全等。

秦湛（1988– ），男，北京人，博士，浙江大学研究员、博士生导师，主要研究方向为数据安全、隐私保护、AI安全等。

陈少梦（1989– ），男，浙江宁波人，杭州快迪科技有限公司工程师，主要研究方向为数据安全、反作弊、反欺诈等。

任奎（1978– ），男，安徽芜湖人，博士，浙江大学教授、博士生导师，主要研究方向为数据安全、人工智能安全等。