# 基于多特征自适应融合的区块链异常交易检测方法

朱会娟[1,2]，陈锦富[1,2]，李致远[1,2]，殷尚男[1,2]
（1. 江苏大学计算机科学与通信工程学院，江苏 镇江 212013；2. 江苏省工业网络安全技术重点实验室，江苏 镇江 212013）

**摘 要**：针对智能检测模型的性能受限于原始数据（特征）表达能力的问题，设计了一种残差网络结构 ResNet-32 用于挖掘区块链交易特征间隐含的关联关系，自动学习包含丰富语义信息的高层抽象特征。虽然浅层特征区分能力弱，但更忠于原始交易细节的描述，如何充分利用两者的优势是提升异常交易检测性能的关键，因此提出了特征融合方法自适应地桥接高层抽象特征与原始特征之间的鸿沟，自动去除其噪声和冗余信息，并挖掘两者的交叉特征信息获得最具区分力的特征。最后，结合以上方法提出区块链异常交易检测模型（BATDet），并通过 Elliptic 数据集验证了所提模型在区块链异常交易检测领域的有效性。

**关键词**：区块链；残差网络；异常检测；Logistic 回归
**中图分类号**：TP18
**文献标识码**：A
**DOI:** 10.11959/j.issn.1000−436x.2021030

# Block-chain abnormal transaction detection method based on adaptive multi-feature fusion

ZHU Huijuan[1,2], CHEN Jinfu[1,2], LI Zhiyuan[1,2], YIN Shangnan[1,2]
1. School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China
2. Jiangsu Key Laboratory of Security Technology for Industrial Cyberspace, Zhenjiang 212013, China

**Abstract:** Aiming at the problem that the performance of intelligent detection models was limited by the representation ability of original data (features), a residual network structure ResNet-32 was designed to automatically mine the intricate association relationship between original features, so as to actively learn the high-level abstract features with rich semantic information. Low-level features were more transaction content descriptive, although their distinguishing ability was weaker than that of the high-level features. How to integrate them together to obtain complementary advantages was the key to improve the detection performance. Therefore, multi feature fusion methods were proposed to bridge the gap between the two kinds of features. Moreover, these fusion methods can automatically remove the noise and redundant information from the integrated features and further absorb the cross information, to acquire the most distinctive features. Finally, block-chain abnormal transaction detection model (BATDet) was proposed based on the above presented methods, and its effectiveness in the abnormal transaction detection is verified.

**Keywords:** block-chain, residual network, abnormal detection, Logistic regression

## 1 引言

科技的飞速发展促使金融行业从实体金融走向互联网金融，反洗钱的外部环境和内在逻辑均发生了深刻而复杂的变化。尤其是随着区块链技术的出现，以比特币为代表的虚拟“数字货币”涌入金

**收稿日期**：2020−08−26；**修回日期**：2020−11−20
**通信作者**：陈锦富，jinfuchen@ujs.edu.cn
**基金项目**：国家重点研发计划基金资助项目（No.2017YFB1400700）；江苏省前沿引领技术基础研究专项基金资助项目（No.BK20202001）；国家自然科学基金资助项目（No.61802154, No.61702230, No.U1836116）
**Foundation Items:** The National Key Research and Development Program of China (No.2017YFB1400700), The Leading-edge Technology Program of Jiangsu Natural Science Foundation (No.BK20202001), The National Natural Science Foundation of China (No.61802154, No.61702230, No.U1836116)

融市场[1-2]。借助比特币实现低成本、点对点的跨境交易成为可能，比特币交易在一定程度上降低了金融行业的准入门槛[3]。比特币属于公有链，其用户参与读取、交易以及共识机制等是开放的，其用户规模是动态的，参与者身份是匿名的[4]，这将导致洗钱犯罪的过程更加隐蔽，洗钱方法也更趋向于复杂化和智能化，从而给传统的反洗钱监管制度体系，如 KYC（know your customer），带来新的问题和挑战。虽然比特币交易场景很难实现 KYC 监管，但是其完整的交易数据在区块链上是公开透明的。顺应反洗钱形势的新变化以及迎接这些公开数据提供的新机遇，一些加密货币情报公司应运而生，例如著名的 Elliptic 公司对链上的交易数据进行挖掘，建立了多维度的数据模型，使利用大数据和人工智能等技术实现数据驱动的智能化反洗钱监管方案成为可能。

该类智能化反洗钱监管方案的首要目标是精准识别区块链上的异常交易，其难点和问题在于：从错综复杂的区块链交易中抽取的原始特征不可避免地存在部分重要关联信息缺失的问题，且受时间、市场行情或恶意攻击等诸多不可控因素影响，原始特征对合法/非法交易的表征能力和区分能力较弱；需要从持续增长的海量交易中准确高效地对少量非法交易进行分类（即正负样本数量极端不平衡）。深度学习凭借其强大的自动特征学习能力给此类问题提供了新的解决方向[5]。现有的一些研究表明，深度学习技术可以模拟人脑的分层模型结构，从原始输入数据中挖掘越来越抽象的且具有良好泛化能力的特征，深度学习模型得到的特征往往具有一定的语义特征和更强的区分能力[6-7]。本文对经典深度学习模型的构造方法、逐层特征变换以及模型训练等方面的关键问题和难点进行研究。例如，深度学习优化过程中隐含层的个数设定目前尚未有明确的规则或规律可循，深层网络较浅层网络具有更强的数据拟合能力，但是层数太深会导致网络退化现象。因此，本文结合残差网络（ResNet, residual network）在解决网络性能退化方面的天然优势，设计了一种残差网络结构 ResNet-32，用来学习交易行为隐含的且具代表性的高层抽象特征。

ResNet-32 提取的高层抽象特征包含了丰富的语义信息，但低层原始特征更倾向于对交易细节本身的描述，如何充分结合两者的优势以进一步提升异常交易检测的性能成为一个迫切需要解决的问题。因此，本文提出了 3 种特征融合方法，即 RRCF（ResNet and raw concat fusion）、RRSF（ResNet and raw supervised fusion）和 RRUF（ResNet and raw unsupervised fusion），以期自适应地桥接高层抽象特征与原始特征，自动去除不同级别特征结合带来的噪声和冗余信息等，并挖掘两者的交叉特征信息以获得最具区分能力的特征，使后续异常交易检测性能的提升成为可能。本文的主要贡献如下。

1) 设计了一种残差网络结构 ResNet-32，用于挖掘孤立特征间错综复杂的关联关系，以期自动学习包含丰富语义信息的高层抽象特征。

2) 提出了一种提升区块链异常交易检测性能的观点，即通过自适应特征融合方法充分挖掘高层抽象特征和原始特征各自的优势。

3) 基于本文提出的 3 种自适应特征融合方法，提出了一种新的区块链异常交易检测（BATDet, det block-chain abnormal transaction detection）模型，并在 Elliptic 数据集[3, 8]上验证了所提模型在区块链异常交易检测方面的有效性。

## 2 相关工作

区块链技术的快速发展促使信息互联网向价值互联网转型，其应用场景广泛，但由于监管机制的缺乏衍生出许多风险或违法违规行为，如洗钱、逃税和非法 ICO 融资等。各国纷纷拟将区块链技术纳入监管体系，其中异常交易检测对区块链产业的健康发展起到了积极的推动作用，例如，余春堂等[9]基于异常交易检测提出了一种分级多层智能服务交易监管架构，是针对具体应用实现的可监管区块链系统。同理，异常交易检测作为反洗钱监管的首要目标，对保证金融市场的稳定发挥着重要作用。在基于 KYC 的传统金融监管体系中，异常交易检测的常规方案是首先设计基于固定阈值规则的警报系统来检测和标记可疑交易，然后对可疑行为进行人工决策或判断[10]。例如 Demetis 等[11]提出了一套适用于银行业的反洗钱系统理论概念模型，以英国一家银行的交易记录作为案例进行研究，提取了一组反洗钱评估指标供反洗钱决策者使用。但是此类监管方案面临的挑战包括：1) 如何从海量异构的交易数据中构建有效规则，并保持规则的先进性和相关性；2) 如何设置洗钱嫌疑交易行为标定的告警阈值。

虚拟“数字货币”等互联网金融的出现导致基于规则的监管方案面临巨大挑战。打破传统的反洗钱监管思维，构建以数据为基础、以人工智能和大数据分析等技术为手段的智能化反洗钱监管方案已成趋势。例如，Weber 等[3]将比特币交易映射为庞大、复杂的图结构，并提取交易数量、交易金额等相关特征，然后采用图卷积网络（GCN, graph convolutional network）算法区分非法和合法交易；Jullum 等[10]利用发送方/接送方的背景、交易早期行为和交易历史等信息训练 XGBoost 有监督预测模型，以识别金融交易中潜在的洗钱行为，并应用于银行。Paula 等[12]从注册信息、金融交易及电子发票等相关类别中提取 18 个重要特征并结合自编码器（AE, auto-encoder）算法训练无监督深度学习模型，以检测和反洗钱相关的出口欺诈。

智能化反洗钱监管中异常交易检测的研究工作目前主要集中在有监督学习和无监督学习[10, 13]。有监督学习通过使用标注数据（训练集）来学习区分的二分类（如合法与非法交易）或多分类机器学习检测模型，从而预测未知数据样本（测试集）的分类。无监督学习则探索未标注数据的结构及特征，找出簇或类的最优划分，将远离其他样本点的孤点视为离群点，即异常数据[14]。但是在反洗钱监管中，异常交易往往隐藏于大量正常交易中且可能刻意模仿正常交易行为[3, 15]，导致合法与非法（异常）交易界限不明显且两者的特征值差异较小，因此无监督方法检测的非法交易不一定是实际场景中的真正非法交易，即存在高误报率和漏报率。

通过以上的分析总结，本文对比分析了目前反洗钱监管领域中异常交易检测相关方法的优势和不足，如表 1 所示。

数据和特征一定程度上决定了智能检测模型的性能上限[16]，特征学习作为提升特征质量的重要手段，近年来伴随深度学习的兴起，实现了从基于领域专家知识的人工构建到从大数据中自主学习的转变[17]。例如，王勇等[18]将网络流量数据进行归一化处理后映射成灰度图片，输入深度卷积神经网络自动提取更加抽象的特征，以提高流量分类的精度；刘烨等[19]探索卷积神经网络、循环神经网络和注意力机制等方法各自的优势来学习软件缺陷报告的重要文本特征和序列特征。得益于深度学习技术的结构优势，基于深度学习的特征融合逐渐被提出，进一步推动了检测领域的发展[20-22]。例如，Pang 等[20]提出的 Libra R-CNN 通过整合不同级别的特征，生成更均衡的语义特征；Zhang 等[23]通过融合预先训练的卷积网络的低层和高层特征来提高语义分割的性能。

针对智能化反洗钱监管方案的现状及发展趋势，本文在特征学习部分充分发挥有监督学习的优势，设计了一种残差网络结构 ResNet-32，用于自动挖掘区块链交易数据中孤立特征间隐含的关联关系，实现对交易数据更本质的刻画；在特征融合部分，本文旨在消除高层特征与低层特征整合后可能存在的信息冲突或噪声，自适应地调整两者特征权重以充分挖掘它们各自优势，从而提升异常交易检测性能。

## 3　区块链异常交易检测模型

### 3.1　模型架构概述

本文提出的 BATDet 模型框架如图 1 所示。考虑经深度学习逐层非线性变换后学习到的高层抽象特征包含更强的语义信息，低层特征（例如原始特征）则侧重于事物本身的描述，本文提出了 3 种特征融合方法，分别是 RRCF、RRSF 和 RRUF。其中，RRCF 将原始交易特征与 ResNet-32 输出的高层抽象特征进行拼接（C, concat），拼接后的特征作为 LR（Logistic regression）分类器的输入。RRSF 在 RRCF 的基础上，将拼接后的特征输入两层堆叠去噪自编码机（SDAE, stacked de-noising auto-encoder）[24]，并利用反向传播（BP, back propagation）算法进行有监督微调。与 RRSF 不同，RRUF 仅采用 SDAE 前向网络，是一种无监督特征融合方法。RRSF 与 RRUF 的目标在于，自适应地调整高层抽象特征与低层原始特征融合时的权重。本文通

**表 1**　　**反洗钱监管领域中异常交易检测方法对比分析**

| 方法 | 优势 | 不足 |
|---|---|---|
| 基于规则 | ①简单、快捷，②适于单一静态场景 | ①难以应对复杂动态场景，②过度依赖人工参与、缺乏灵活性 |
| 无监督学习 | ①不需要标签，②自动化程度高 | ①高误报率与高漏报率，②对特征中的噪声敏感 |
| 有监督学习 | ①预测精度高，②适用性广，③结果容易理解 | ①需要高质量标记样本，②检测性能受限于原始数据（特征）的表达能力 |

过 Ellipict 数据集验证了这 3 种特征融合方法提升 BATDet 模型性能的有效性。

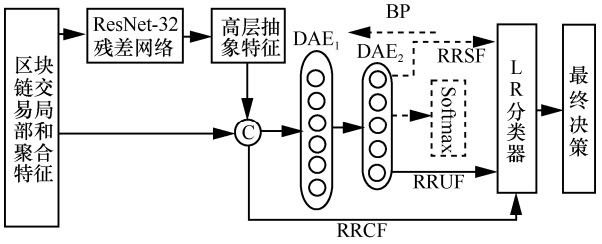


图 1 BATDet 模型框架

### 3.2 Elliptic 数据集

Elliptic 数据集[3]将原始比特币（BTC, bitcoin）交易信息映射为图结构，节点表示比特币交易，边代表不同交易间的比特币流向。它共包含 203 769 个节点和 234 355 条边，节点上同时携带时间戳信息。在 Elliptic 数据集里，比特币交易主体被纳入合法实体（如交易所、钱包提供商、合法服务等）和非法实体（如诈骗、恶意软件、勒索软件等）。合法实体发起的交易被标记为合法交易，非法实体发起的交易被标记为非法交易。采用启发式的推理过程，该数据集共标注了 4 545 笔非法交易和 42 019 笔合法交易，其余交易的标记未知。依据节点的时间戳信息，这些已标记的交易被划分为 49 个独立的时间片。每笔交易包含 166 个特征，其中 94 个特征是比特币交易（节点）的本地信息，记为局部特征，如时间步长、交易手续费、输入/输出数、输出量和一些合计数据（如平均接收/转出的 BTC 数量等）；其余 72 个特征是从中心节点向前/向后一跳聚合事务信息统计出的最大值、最小值、标准差和相关系数等，标记为聚合特征。

### 3.3 特征学习

针对比特币交易特征间存在高度非线性关联的特点，本文采用深度学习算法进行特征学习和融合。深度学习作为人工智能的一个重要分支，与浅层学习方法相比，能够自动地学习复杂非线性变换函数，提取更抽象和有效的隐含特征[5, 25]，使原始数据信息利用率大幅提升。这些包含数据丰富内在规律的深层特征已在诸多领域取得显著成效，例如目标分类检测、语义分割和行为识别等领域[26]。但是大多数深度学习方法往往面临参数多、训练难度大、梯度消失和网络退化等问题[27-28]，从而影响或限制后续检测或分类任务的性能。深度学习常规架构包含输入层、隐含层和输出层，在硬件环境和原始数据许可的情况下深度学习往往通过加深隐含层数来提升特征学习能力。但是，大量实验证明当层数到达一定深度（即最优结构）后，性能反而会随层数的增加而下降（并非由过拟合导致，可能是梯度消失等原因导致），即网络退化问题[28-29]。在深度学习优化过程中隐含层的层数设置目前尚未有明确的规则或规律可循。

He 等[28]提出的 ResNet 通过引入恒等映射来解决网络退化问题。ResNet 中最基本单元是残差块。每个残差块均包含 2 个具有相同核大小和过滤数的卷积，将其输入记为 $x$，输出（即学习到的特征）记为 $H(x)$，残差网络中通过叠加非线性层来拟合一个残差函数 $F(x):=H(x)-x$。假设网络结构（共 $l$ 层）达到 $n$ 层（$n<l$）时为最优结构，则 $n\sim l$ 层为冗余层，也就是恒等映射即为最优。残差网络可将残差 $F(x)$ 逼近 0 来完成恒等映射，理论上，$l$ 可以继续增加，且准确率不会下降，从而解决了网络退化问题。因此，本文采用 ResNet 自动学习区块链交易中隐含的且具代表性的高层抽象特征。

为了获得包含更多语义信息的深层次特征，交易的特征提取网络通过级联如图 2 所示的残差块构成。

每个残差块的输出可定义为

$$H(x)=F(x,\{W_i\})+x \tag{1}$$

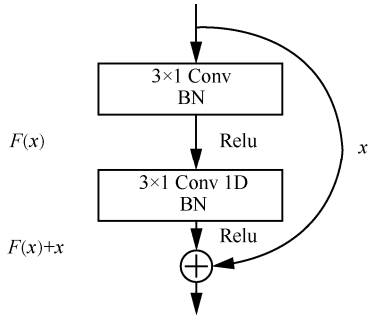


图 2 残差块

在本文的 ResNet-32 中，第一层输入为交易的原始特征，共 166 维；$F(x,\{W_i\})$ 为需要学习的残差映射。如图 2 所示，本文采用的跳跃连接设置为 2，可得

$$F(x,\{W_i\})=\delta(W_2\delta(W_1x)) \tag{2}$$

其中，$\delta$ 表示非线性激活函数 Relu[30]，因为相比 sigmoid 或 tanh 激活函数，它具有更快的收敛速度和更好的性能[17, 31]。恒等映射通过跳跃连接完成，

并与指定输出层做加法操作，此时需要 $x$ 和 $H(x)$ 形状（包括维度等）相同，如果不相同，则式(1)转换为

$$H(x)=F(x,\{W_i\})+W_s x \tag{3}$$

其中，$W_s$ 采用1×1卷积操作实现，用于调整 $x$ 的维度。另外，在每一层卷积之后引入了批归一化（BN, batch normalization），以避免梯度消失和加快收敛[32]。用于交易数据深层次特征学习的残差网络结构如图 3 所示，对应的参数设置如表 2 所示。

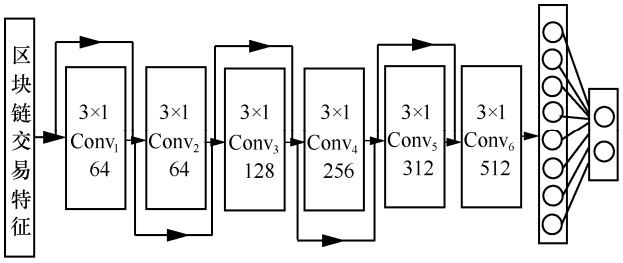


图 3　用于交易数据深层次特征学习的残差网络结构

**表 2　　残差网络参数设置**

| 块名 | 输出形状 | 块结构（32 层） |
|---|---|---|
| $Conv_1$ | 166×1 | 3x1,64,stride 2<br>BN 1D（64）,MaxPool,3×1,stride 2 |
| $Conv_2$ | 166×1 | $\begin{bmatrix}3\times1,64\\3\times1,64\end{bmatrix}\times3$ |
| $Conv_3$ | 83×1 | $\begin{bmatrix}3\times1,128\\3\times1,128\end{bmatrix}\times3$ |
| $Conv_4$ | 42×1 | $\begin{bmatrix}3\times1,256\\3\times1,256\end{bmatrix}\times3$ |
| $Conv_5$ | 21×1 | $\begin{bmatrix}3\times1,312\\3\times1,312\end{bmatrix}\times3$ |
| $Conv_6$ | 11×1 | $\begin{bmatrix}3\times1,512\\3\times1,512\end{bmatrix}\times3$ |
| 全连接层 | 1×1 | 全局平均池化 512-2 FC, Softmax |

### 3.4　多特征融合

残差网络学习到的深层次抽象特征经过逐层的非线性变换及有监督微调，不再孤立地考虑交易信息中的各个特征，而是有效捕获表征交易信息的各个特征间错综复杂的内在非性关联关系，挖掘其蕴含的语义信息。与深层次抽象特征包含丰富语义信息不同，浅层次特征（例如原始特征）更忠于事物本身的描述，例如交易的细节信息等。如何桥接这 2 种特征并挖掘两者优势，以增强特征区分能力与抗噪性能，是异常交易检测面临的另一挑战。

针对上述问题，本文提出了自适应特征融合方法，以探索高层抽象特征与浅层原始特征的不同级别信息之间的优势互补。首先，本文提出 RRCF，即通过连接操作将残存网络输出的深层次特征与原始特征映射到统一的特征表示空间，记为 $x_{\text{inputf}}$，如式(4)所示。

$$x_{\text{inputf}}=[\text{raw\_features},\text{resnet32\_features}]^{\text{T}} \tag{4}$$

但是，拼接后的特征面临如下问题：1) 噪声数据，即原始特征 raw_features 与残差网络学习到的特征 resnet32_features 这 2 组不同层次的特征拼接后，在后续分类阶段可能存在干扰、冲突或不一致信息，从而产生数据噪声；2) 如何界定这 2 组特征各自的重要程度，即相应的权重如何分配才能获取更具有区分能力的高品质特征。

鉴于 SDAE 逐层贪婪训练及从噪声数据中学习稳健性和强泛化的紧凑不变特征的能力，在特征融合阶段，本文设计了基于 SDAE 的特征融合方法 RRSF 和 RRUF，以自动学习 $x_{\text{inputf}}$ 中不同特征的权重，并有效抑制特征间的冲突和不一致等噪声信息。RRSF 和 RRUF 的差异在于，RRSF 中加入了 Softmax 层利用标签对其进行有监督微调，RRUF 采用的是无监督前向网络。SDAE 是通过堆叠多个 DAE（DAE, de-noising auto-encoder）构成的。DAE 是一种三层神经网络模型（包含输入层、隐含层和输出层），给定输入数据 $x_f$（输入层），随机将 $x_f$ 中的部分元素置 0 以获得噪声数据 $\tilde{x}_f$，则隐含层 $h_f$ 表示为

$$h_f=\boldsymbol{F}_{W,b}(\tilde{x}_f)=\varsigma(\boldsymbol{W}\tilde{x}_f+b) \tag{5}$$

其中，$\varsigma$ 是非线性激活函数 LeakyReLU，$\boldsymbol{W}$ 为编码器的权重矩阵，$b$ 为偏差，$h_f$ 的维度小于 $x_f$ 的维度。在解码器部分对 $h_f$ 进行线性或非线性变换，获得输出层 $x'_f$ 为

$$x'_f=\boldsymbol{F}_{W',b'}(h_f)=\boldsymbol{W}'h_f+b'\text{or}\varsigma'\left(\boldsymbol{W}'h_f+b'\right) \tag{6}$$

其中，$\boldsymbol{W}'$ 为隐含层到输出层的权值矩阵，$b'$ 为偏差，$\varsigma'$ 为非线性激活函数。在 BATDet 模型中，为了加快训练速度，融合阶段解码器采用线性变换。计算输入层 $x_f$ 与输出层 $x'_f$ 之间的误差，通过误差反向传播算法不断调整权值和偏差 $\boldsymbol{W}$、$b$、$\boldsymbol{W}'$、$b'$，使重构误差最小，即转化为解决式(7)所示问题。

$$\min \sum \left\| x'_f - x_f \right\| \tag{7}$$

本文设计的 SDAE 为二层结构，即 2 个 DAE 的级联，融合后的特征 $x_{\text{outputf}}$ 可表示为

$$x_{\text{outputf}}=\varsigma\left(\boldsymbol{W}_2\varsigma(\boldsymbol{W}_1 x_{\text{inputf}}+b_1)+b_2\right) \tag{8}$$

针对传统特征融合方法无法伴随特征集的变化进行动态优化的问题，BATDet 模型集深层抽象特征与浅层原始特征各自优势于一体，提出了新的区块链交易自动特征学习和融合方法，以自动捕获交易行为全方位的且具代表性的特征，该方法具有一定的通用性和普适性，可移植和扩展到其他应用领域。

## 4 性能测试与分析

本节基于 Elliptic 数据集对 BATDet 模型的性能进行综合评估和分析。考虑区块链异常交易检测是一个与时间信息密切相关的任务，且在 Elliptic 数据集中将已标注的交易记录依据其发生的时间点划分为 49 个独立的时间片。为了尽可能保持数据分布的一致性，本文采用留出法，依据时间片比例 70:30 划分训练集和测试集，即前 34 个时间片的交易记录用于训练特征学习、融合以及后续的分类模型，第 35 个～第 49 个时间片的交易记录将用于测试和评估模型的性能。

首先，采用经典的 LR 分类器评估 Elliptic 数据集中 166 维原始特征（包括比特币交易的局部特征和近邻聚合特征）对正常交易和异常交易的区分能力，以其作为基准，评估了本文设计的 ResNet-32 残差网络的特征学习能力和所提不同层次的多特征自适应融合方法对区块链交易异常检测性能的提升；其次，针对区块链交易具有较强时效性的特点，进一步评估了 BATDet 模型在新发生异常交易检测时的稳健性；最后，在同等实验条件下，将 BATDet 模型与经典智能检测模型，如 k 最近邻（KNN, k-nearest neighbor）、决策树（DT, decision tree）、多层感知机（MLP, multilayer perceptron）、朴素贝叶斯（NB, naive Bayesian）、Adaboost 和梯度提升树（GBDT, gradient boosting decision tree）进行对比，以评估其整体有效性。BATDet 模型中各特征学习与融合方法均采用 LR 分类器进行验证。

### 4.1 评估标准

为了客观评估和度量 BATDet 模型在区块链异常交易检测方面的性能，考虑本文区块链异常交易检测的目的是区分正常交易和异常交易，属于二分类问题。因此，本文引入机器学习领域经典的二分类评估指标，包括准确率 Accuracy、精确率 Precision、召回率 Recall、F1 值 $F1$-score、马修斯相关系数（MCC, Matthews correlation coefficient）等，其详细定义如下

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{9}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

$$F1-\text{score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

$$\text{MCC}=\frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{13}$$

其中，TP（true positive）、FP（false positive）、TN（true negative）和 FN（false negative）分别表示真正例数、假正例数、真反例数和假反例数。

### 4.2 实现及参数设置

本文在 Pytorch 深度学习框架下实现了 BATDet 模型，具体的参数设置如下。1) 高层抽象特征通过 ResNet-32 获取，网络结构参数如表 2 所示，epoch 设置为 200（读取训练集中全部样本完成一次训练即为 1epoch），学习率为 0.000 1。2) 高层抽象特征与底层特征融合阶段，级联的 2 个 DAE 的隐含层神经元个数分别为 416 和 256，epoch 为 100，学习率为 0.000 1，为了加快特征融合阶段的训练过程，本文将训练集分割成小批量来更新模型参数，其中，批大小为 300，epoch 为 100。BATDet 模型中均采用 Adam 优化器[33]更新网络参数。另外，本文采用基于 Python 的机器学习算法库 Sklearn0.19.0 实现 LR、KNN、DT、NB、Adaboost 和 GBDT。

### 4.3 性能评估及比较

为了直观展示 BATDet 模型中各组成部分对区块链异常交易检测性能的影响，本节首先将 BATDet 模型分解为如下 5 个步骤进行评估，检测结果如表 3 所示。

1) 评估 Elliptic 数据集中 166 维原始特征在 LR 分类器上的异常交易检测性能，记为 Raw。

2) 测试 BATDet 模型中设计的 ResNet-32 网络结构提取的高层抽象特征对异常交易检测性能的提升，记为 ResNet-32。

3) 验证高层抽象特征与底层原始特征经简单拼接后的检测性能，即 RRCF。

4) 评测高层抽象特征与底层原始特征经有监督微调的 SDAE 进行融合后的异常交易检测性能，即 RRSF。

5) 评估经无监督 SDAE 进行融合特征权重学习后的检测性能，即 RRUF。

**表 3　　区块链异常交易检测结果**

| 方法 | Accuracy | Precision | Recall | *F*1-score | MCC |
|---|---|---|---|---|---|
| Raw | 88.70% | 32.86% | 70.91% | 69.31% | 43.20% |
| ResNet-32 | 97.46% | 86.18% | 72.58% | 88.72% | 77.79% |
| RRCF | 97.53% | 87.49% | 72.30% | 88.93% | 78.27% |
| RRSF | 97.77% | 92.89% | 71.20% | 89.71% | 80.24% |
| RRUF | 97.86% | 94.59% | 71.10% | 90.02% | 80.99% |

首先，利用 Elliptic 数据集中 166 维原始特征训练 LR 分类器，以评估原始特征在异常交易检测中的表现，以此作为基准评估本文所提出的 ResNet、RRCF、RRSF 和 RRUF 的有效性。从表 3 可知，在相同的实验环境及参数设置下，本文设计的 ResNet-32 学习到的特征作为 LR 分类器的输入特征可以显著提高异常交易检测性能。较基准方法 Raw，Accuracy 提升了 8.76%，Precision 提升了 53.32%，Recall 提升了 1.67%，*F*1-score 提升了 19.41%，MCC 提升了 34.59%。ResNet-32 带来的检测性能大幅度提升的主要原因在于：1) 残差网络通过引入恒等映射等概念有效解决了网络退化问题，使优化较深层的网络以提升特征学习能力成为可能；2) ResNet-32 学习到的深层非线性网络结构可以有效拟合输入数据，以组合和关联区块链交易中各孤立特征，并逐层进行非线性特征变换，展现了从训练样本集中学习区块链交易高级语义特征的强大能力，从而使异常交易检测性能得以大幅度提升。

然而，特征变换的过程中不可避免地会丢失一些交易的细节信息，低层原始特征虽然稀疏且可能含有噪声或语义歧义干扰，导致其区分能力较弱，但它是对交易细节的最直观表述。因此，为了充分挖掘两者的互补优势，本文提出了 3 种特征融合方法。其中，RRCF 为底层原始特征与高层语义特征的拼接。通过表 3 的结果可知，相较于 ResNet-32，除召回率以外，RRCF 的各项评价指标均有提升，实验结果进一步说明融合底层与高层特征可以带来检测性能的提升。但不同级别特征简单的拼接，可能会带来信息冗余或噪声，且很难界定它们在异常检测场景中的重要程度。因此，本文结合 SDAE（本文采用级联两层 DAE）从噪声数据中学习稳健性和强泛化特征的优势，进一步提出了 2 种特征融合方法 RRSF 和 RRUF。在 RRCF 的基础上，RRSF 与 RRUF 通过缩小输入与输出误差自动调整对应的网络权重，最大限度地保留不同层次特征各自的差异和优势，并挖掘两者的交差信息，以获得更具区分能力的特征。

Elliptic 数据集中非法交易（即异常交易）数与合法交易数的比值为 4 545：42 019，这种样本分类极度不平衡的情况在异常交易检测场景中是普遍现象。由表 3 中实验结果可知，在特征融合阶段，与较简单的 RRCF 特征拼接方法相比，RRSF 和 RRUF 的 Precision 分别提升了 5.4%和 7.1%（不平衡数据集的重要参考指标）。由此可见，本文提出的自适应特征融合方法可有效提升区块链异常交易检测性能。其中，RRUF 优于 RRSF 主要原因在于高层抽象特征学习与融合特征学习阶段，由于训练数据相同，因此它们均采用了相同的标签进行有监督微调，故融合阶段学习到的特征在重要性上会更倾向于高层抽象特征而弱化底层特征。

为了评估 BATDet 模型对新发生异常交易检测的有效性，本文采用时间片 1～时间片 34 的交易记录为训练集，分别评估了时间片 35～时间片 49 中发生的交易，如图 4～图 8 所示。从 Accuracy、Precision、Recall、*F*1-score 和 MCC 这 5 个指标分别对 BATDet 模型中的检测方法进行评估和比较。从实验结果来看，本文提出的 BATDet 模型中的 ResNet-32 以及 3 种特征融合方法 RRCF、RRSF 和 RRUF 在时间片 35～时间片 42 中，极大地提升了区块链异常交易检测性能中除 Recall 外的其他指标。Precision 和 Recall 是一对矛盾的度量标准，通常 Precision 高时，Recall 会偏低。

从图 5～图 8 的对比可以看出，在时间片 43 及其后的部分时间片中，BATDet 模型中所提方法并未使异常交易检测性能明显提升，这是由于在时间片 43 发生了黑市突然关闭现象[3]，即在这个时间片的共 1 370 笔交易中只有 24 笔交易被标注为非法交易，从而导致所有检测方法均无法捕获黑市关闭后发生的非法交易，这是异常交易检测在反洗钱应用领域需应对的一个重要挑战。本文认为在时间片 43～时间片 49 中，所有模型的检测性能均欠理想，

极大可能是随着时间的推移及交易数据分布的变换，导致预训练的模型不适用于新发生交易的检测，即发生了概念漂移现象。针对此类问题，本文下一步工作拟结合深度学习优异的特征学习能力，以及迁移学习在解决数据分布差异及标注数据过期等问题上的优势展开进一步的研究。

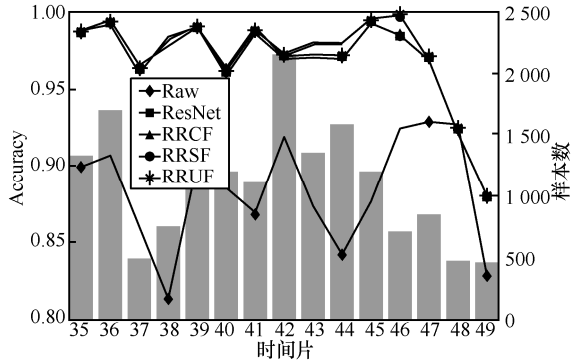


图 4 不同时间片中异常检测的 Accuracy 比较

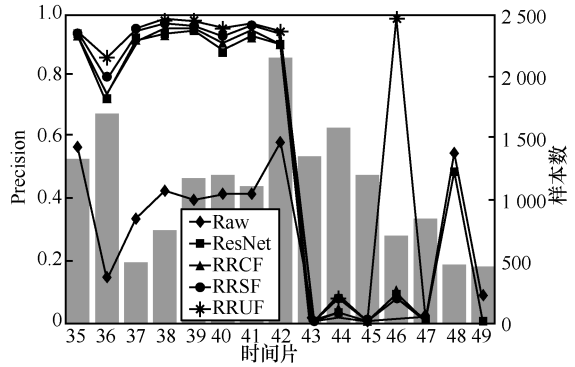


图 5 不同时间片中异常检测的 Precision 比较

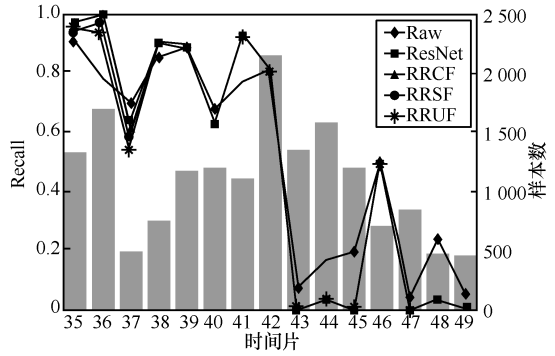


图 6 不同时间片中异常检测的 Recall 比较

为验证所提 BATDet 模型在区块链异常交易检测上的整体性能，本文在 Elliptic 数据集上选择与经典的分类模型如 KNN、DT、MLP、NB、Adaboost 和 GBDT 进行比较，结果如表 4 所示。实现方式均采用 Python 语言实现的机器学习算法库 Sklearn0.19.0，所有分类器均采用默认参数。

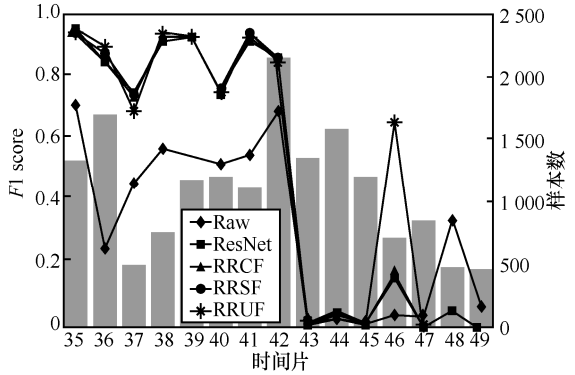


图 7 不同时间片中异常检测的 *F*1-score 比较

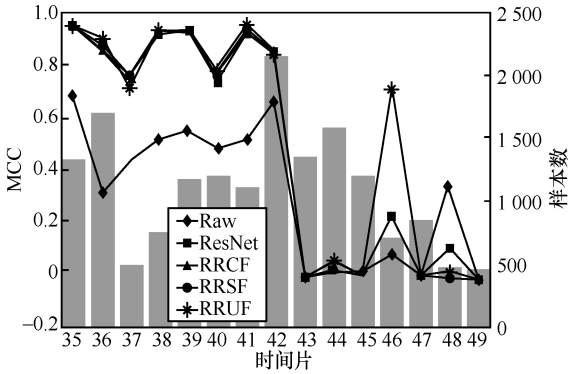


图 8 不同时间片中异常检测的 MCC 比较

**表 4 与经典智能检测模型的比较**

| 模型 | Accuracy | Precision | Recall | *F*1-score | MCC |
|---|---|---|---|---|---|
| KNN | 93.66% | 50.95% | 64.45% | 76.74% | 53.97% |
| DT | 92.47% | 45.28% | 76.18% | 76.34% | 55.12% |
| MLP | 95.13% | 63.62% | 59.46% | 79.31% | 58.77% |
| NB | 70.16% | 16.33% | 87.17% | 54.36% | 29.07% |
| Adaboost | 96.21% | 72.29% | 67.68% | 83.94% | 67.93% |
| GBDT | 97.10% | 80.72% | 72.67% | 87.47% | 75.06% |
| ResNet-32 | 97.46% | 86.18% | 72.58% | 88.72% | 77.79% |
| RRCF | 97.53% | 87.49% | 72.30% | 88.93% | 78.27% |
| RRSF | 97.77% | 92.89% | 71.20% | 89.71% | 80.24% |
| RRUF | 97.86% | 94.59% | 71.10% | 90.02% | 80.99% |

从表 4 可以看出，在正负样本数量极度失衡的现实情况下，本文的特征学习方法 ResNet-32 和特征融合方法 RRCF、RRSF 和 RRUF 获得的特征更适用于区块链异常交易检测，其 Precision 大大提高。其他指标如 Accuracy、Recall、*F*1-score 和 MCC 也得以大幅度提升。实验结果进一步表明 BATDet 模型可以有效拟合比特币交易数据，通过挖掘特征间关联关系以及自适合融合多级别特征来提高原始数据信息利用率，保证了在数据极度不平衡的条

件下异常交易检测的可靠性和稳定性。

## 5　结束语

伴随区块链等新兴技术的兴起，基于虚拟“数字货币”的在线交易降低金融行业的准入门槛和提供便捷金融服务的同时，也因其匿名性等特点为不法分子提供了更隐秘的洗钱途径，这给目前金融行业通用的基于 KYC 的反洗钱监管机制带来了极大的挑战。区块链上数据的公开透明和不可篡改等特性，结合人工智能和大数据分析及挖掘等技术，为反洗钱监管提供了新的契机。本文首先以区块链异常交易检测为切入点，提出深度残差网络结构 ResNet-32 学习交易特征间的非线性关联关系，以解决受时间、市场行情或恶意攻击等诸多不可控因素影响而导致的交易原始特征表征能力和区分能力较弱的问题；其次，提出了自适应特征融合方法 RRCF、RRSF 和 RRUF，以提升信息利用率为目的桥接高层抽象特征与原始特征和挖掘两者优势，进一步提升区块链异常交易检测的性能；最后，通过 Elliptic 数据集分步骤评估了本文提出的 BATDet 模型的可靠性和有效性，表明本文方法可为虚拟“数字货币”背景下的反洗钱监管提供新的解决方案。

## 参考文献：


[1] 龚鸣. 非基于 KYC 的区块链反洗钱技术探讨[J].清华金融评论, 2017, 4: 52-54.
GONG M. Discussion on anti money laundering technology of blockchain nor based on KYC[J]. Tsinghua Financial Review, 2017, 4: 52-54.

[2] QI T T, WANG T M, ZHU J M, et al. Empirical study on the correlation and volatility between bitcoin and the blockchain index: based on granger causality test and GARCH-class model[C]//The 4th International Conference on Crowd Science and Engineering. New York: ACM Press, 2019: 28-32.

[3] WEBER M, DOMENICONI G, CHEN J, et al. Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics[J]. arXiv Preprint, arXiv: 1908.02591, 2019.

[4] 钱卫宁, 邵奇峰, 朱燕超, 等. 区块链与可信数据管理: 问题与方法[J]. 软件学报, 2018, 29(1): 150-159.
QIAN W N, SHAO Q F, ZHU Y C, et al. Research problems and methods in blockchain and trusted data management[J]. Journal of Software, 2018, 29(1): 150-159.

[5] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature,2015, 521(7553): 436-444.

[6] REN K, ZHENG T H, QIN Z, et al. Adversarial attacks and defenses in deep learning[J]. Engineering, 2020, 6(3): 346-360

[7] ZHAO Z Q, ZHENG P, XU S T, et al. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.

[8] Elliptic. Elliptic company[R]. (2019)[2020-11-20].

[9] 余春堂, 韩志耕, 李致远, 等. 基于区块链的众包物流分级多层智能服务交易监管架构[J]. 网络与信息安全学报, 2020, 6(3): 50-58.
YU C T, HAN Z G, LI Z Y, et al. Blockchain-based hierarchical and multi-level smart service transaction supervision framework for crowdsourcing logistics[J]. Chinese Journal of Network and Information Security, 2020, 6(3): 50-58.

[10] JULLUM M, LLAND A, HUSEBY R B, et al. Detecting money laundering transactions with machine learning[J]. Journal of Money Laundering Control, 2020, 23(1): 173-186.

[11] DEMETIS D S. Fighting money laundering with technology: a case study of Bank X in the UK[J]. Decision Support Systems, 2018, 105: 96-107.

[12] PAULA E L, LADEIRA M, CARVALHO R N, et al. Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering[C]//2016 15th IEEE International Conference on Machine Learning and Applications. Piscataway: IEEE Press, 2016: 954-960.

[13] CHEN Z Y, KHOA L D, TEOH E N, et al. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review[J]. Knowledge and Information Systems, 2018, 57(2): 245-285.

[14] PHAM T, LEE S. Anomaly detection in the bitcoin system - a network perspective[J]. arXiv Preprint, arXiv:1611.03942v2, 2016.

[15] LORENZ J, SILVA M L, APARICIO D, et al. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity[J]. arXiv Preprint, arXiv: 2005.14635, 2020.

[16] 李韵, 黄辰林, 王中锋, 等. 基于机器学习的软件漏洞挖掘方法综述[J]. 软件学报, 2020, 31(7): 2040-2061.
LI Y, HUANG C L, WANG Z F, et al. Survey of software vulnerability mining methods based on machine learning[J]. Journal of Software, 2020, 31(7): 2040-2061.

[17] LI X, DING M L, PIŽURICA A. Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(4): 2615-2629.

[18] 王勇, 周慧怡, 俸皓, 等. 基于深度卷积神经网络的网络流量分类方法[J]. 通信学报, 2018, 39(1): 14-23.
WANG Y, ZHOU H Y, FENG H, et al. Network traffic classification method basing on CNN[J]. Journal on Communications, 2018, 39(1): 14-23.

[19] 刘烨, 黄金筱, 马于涛. 基于混合神经网络和注意力机制的软件缺陷自动分派方法[J]. 计算机研究与发展, 2020, 57(3): 461-473.
LIU Y, HUANG J X, MA Y T. An automatic method using hybrid neural networks and attention mechanism for software bug triaging[J]. Journal of Computer Research and Development, 2020, 57(3): 461-473.

[20] PANG J M, CHEN K, SHI J P, et al. Libra R-CNN: towards balanced learning for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 821-830.

[21] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[J]. arXiv Preprint, arXiv:1911.09516, 2019.

[22] GHIASI G, LIN T Y, LE Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7029-7038.

[23] ZHANG Z, ZHANG X, PENG C, et al. ExFuse: enhancing feature fusion for semantic segmentation[C]//European Conference on Computer Vision. Berlin: Springer, 2018:273-288.
[24] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research,2010, 11(12): 3371-3408.
[25] LI K, ZOU C Q, BU S H, et al. Multi-modal feature fusion for geographic image annotation[J]. Pattern Recognition, 2018, 73: 1-14.
[26] LI G H, MATTHIAS M, QIAN G C, et al. DeepGCNs: making GCNs go as deep as CNNs[J]. arXiv Preprint, arXiv:1904.03751, 2019.
[27] 刘树东, 王晓敏, 张艳. 一种对称残差 CNN 的图像超分辨率重建方法[J]. 西安电子科技大学学报, 2019, 46(5): 15-23.
LIU S D, WANG X M, ZHANG Y. Symmetric residual convolution neural networks for the image super-resolution reconstruction[J]. Journal of Xidian University, 2019, 46(5): 15-23.
[28] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
[29] 裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测方法[J]. 软件学报, 2019, 30(3): 738-758.
PEI W, XU Y M, ZHU Y Y, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019, 30(3): 738-758.
[30] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines Vinod Nair[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2010: 807-814.
[31] CHEN Y S, JIANG H L, LI C Y, et al. Deep feature extraction and classification of hyper spectral images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10): 6232-6251.
[32] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. arXiv Preprint, arXiv:1502.03167, 2015.
[33] KINGMAD, BA J. Adam: a method for stochastic optimization[C]// International Conference on Learning Representations. Piscataway: IEEE Press, 2015:1-11.

[作者简介]

**朱会娟**（1984− ），女，河南洛阳人，博士，江苏大学副教授、硕士生导师，主要研究方向为区块链、网络安全及人工智能。

**陈锦富**（1978− ），男，江西信丰人，博士，江苏大学教授、博士生导师，主要研究方向为可信软件及软件安全。

**李致远**（1981− ），男，河南开封人，博士，江苏大学副教授、硕士生导师，主要研究方向为物联网、大数据安全。

**殷尚男**（1989− ），男，吉林白城人，博士，江苏大学副教授，主要研究方向为入侵检测及孤立点检测。