基于网络流量的用户网络行为被害性分析模型

周胜利，徐啸炀
（浙江警察学院，浙江 杭州 310051）

摘　要：网络行为被害性分析对于电信网络诈骗犯罪的防控具有深远意义。通过研究用户与网站交互产生的网络流量，提出一种基于网络流量分析的电信网络诈骗犯罪用户网络行为被害性识别模型，分析不同网络行为特征之间的关联规则，重构网络行为序列特征，同时结合随机森林算法评估网络行为的被害性。在被害人网络行为数据集基础上进行实验，证明模型能够有效提升网络行为被害性识别准确率。

关键词：网络流量；网络行为编码；关联规则挖掘；被害性分析
中图分类号：TP311
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2021041

Victimization analysis model of user network behavior based on network traffic

ZHOU Shengli, XU Xiaoyang
Zhejiang Police College, Hangzhou 310051, China

Abstract: The analysis of network victimization is of great significance to the prevention and control of telecom fraud. By studying the network traffic generated by the interaction between users and websites, a victimization identification model of telecom fraud crime based on network behavior flow analysis was proposed, the association rules between different behavior characteristics were analyzed, the behavior sequence features were reconstructed, and the victimization of network behavior sequence with random forest algorithm was evaluated. Based on the network behavior data set of public security organs, the experiment proves that the model can effectively improve the recognition accuracy of network behavior victimization.

Key words: network traffic, network behavior coding, association rules mining, victimization analysis

1　引言

随着电信网络技术和互联网金融业务的快速发展，电信网络诈骗正逐步取代传统诈骗犯罪，成为当前犯罪的主要形式，严重威胁人民生命财产安全。公安部统计数据显示，全国电信网

收稿日期：2020−12−21；修回日期：2021−02−11
通信作者：徐啸炀，76933768@qq.com
基金项目：浙江省公益技术研究计划（No.LGF20G030001）；校局合作项目（No.2020XJY011）；国家级创新项目（No.11483）
Foundation Items: The Basic Public Welfare Research Program of Zhejiang Province of China (No.LGF20G030001), School Bureau Cooperation Project (No.2020XJY011), The National Innovation Project national Innovation Project(No. 11483)

络诈骗犯罪从 2011 年的 84 514 起飙升至 2019 年的 808 730 起；受骗金额从 2011 年的 100 亿元飙升至 2019 年的 192 亿元。尽管当前国家采取了形式多样的预警和防控措施，取得了一定成效，但仍旧缺乏精准、高效的智能监测预警方法，打击防范犯罪形势依然非常严峻。开展基于网络行为流量分析的电信网络诈骗犯罪被害性识别技术研究能够有效提高犯罪预警的精准性和效率，从而更好地保护人民生命财产安全。

针对电信网络诈骗犯罪防控研究中网络行为特征选取表面化、特征间内在规则挖掘不足、网络行为稀疏、行为序列间关系难以确定等问题，本文提出基于网络行为流量分析的电信网络诈骗犯罪被害性识别模型（victimization identification model of telecom fraud crime based on network behavior traffic，VIM-TFCN），分析用户行为特征，挖掘潜在关联规则，综合评估用户被害风险，达到电信网络诈骗犯罪高效预警的目的。

2 相关研究

国内外与本文相关的研究主要集中在电信网络诈骗犯罪识别与防控、网络流量分析、网络行为分析。

2.1 电信网络诈骗识别与防控

电信网络诈骗犯罪研究领域，国内外学者主要利用数据挖掘、自然语言处理等方法进行电信网络诈骗犯罪特征态势[1]、异常通信分析模型[2-3]、语音识别模型[4]等方面研究以及采用网页相似度分析[5]、基于网页关系检测与网站链接评估的检测方法[6]、恶意域名检测方法[7-8]与 BERT 迁移学习方法[9]进行电信网络诈骗平台识别预警研究。以上方法主要针对电信网络诈骗犯罪平台或者诈骗通信进行识别，缺少对电信网络诈骗犯罪被害人网络行为分析预测，且在具体研究方法上存在特征冗余程度较高、识别维度单一等问题。电信网络诈骗防控实战应用领域，Endgame 公司开发网络平台实时分析可疑网络活动，为电信网络诈骗案件侦破助力。360 公司在“通用算法引擎”与“定制化算法引擎”机制构建上取得重大进展，开发智控•商业反欺诈平台实现黑/灰产业链的动态监视。阿里巴巴公司为移动保障安全，开发安全钱盾反诈平台。

2.2 网络流量分析

对于异常流量检测研究，Zolotukhin 等[10]以流量日志分析为基础，提出了一种对攻击 Web 应用行为的异常检测方法。Yu 等[11]、Yang 等[12]等采用自然语言识别的方法，通过建立相关词库进行分词预处理，最终以神经网络模型进行异常检测。Park 等[13]提出了基于二值图变换的卷积自动编码器，对流量数据分组进行异常检测。在提取流量的有效数据上，Arzhakov 等[14]提出使用蜜罐技术收集用户行为统计信息，并基于统计结果来区分不同种类的流量。Thang 等[15]建立了基于密度的噪声应用空间聚类模型来提取流量中的有效数据。在混合模型方面，Zhang 等[16-17]提出采用隐马尔可夫模型、概率分布模型、支持向量机等模型对 HTTP 请求进行异常检测。

2.3 网络行为分析

当前对网络行为分析的研究主要包括网络行为异常行为检测和推荐系统研究。

异常行为检测方面，连一峰等[18]采用关联分析与序列挖掘技术，通过比较用户当前行为模式与历史行为模式的相似度判断异常。该方法能够实现对用户异常行为的检测，但缺少应对大规模数据的能力。田新广等[19]针对上述模型的不足，改进了用户行为模式的表示方式，联合采用多个判决门限对用户行为进行判断，并提出 IDS 异常检测模型。该模型具备更高的检测效能，但存在应用范围有限和检测边界模糊的问题。陈胜等[20]为了解决传统异常行为检测方法难以应对海量数据，无法及时响应新行为的问题，提出了一种基于深度神经网络，并能够自定义用户行为的检测模型。该模型拥有海量

数据检测能力，能够检测未知的异常行为，具有较高精确度与鲁棒性。胡富增等[21]探究用户行为特征及行为模式，采用聚类分析算法，对用户日志数据进行数据挖掘与聚类分析，最终实现行为模型的建立。该方法具有简便的特点，但不足之处在于识别成功率偏低，准确率不足。

推荐系统研究方面，Wang 等[22]通过比较其他传统的推荐系统（recommender system，RS），提出了一种基于会话的推荐系统（session-based recommender system，SBRS）模型分析用户行为，并在此基础上引入了等级分层框架，分析讨论了推荐模型在用户行为分析上的优点与不足。Tang 等[23]建立了一种卷积嵌入的 Top-*N* 序列推荐模型，采用卷积滤波器进行分析，具有较理想的效果。Sun 等[24]同样使用卷积神经网络进行建模，提出一种双向编码的序列推荐模型。该模型对用户行为采用双向自注意机制，解决了当下推荐模型中普遍存在的用户动态取向及历史行为问题。Hidasi 等[25]采用了循环神经网络（recurrent neural network，RNN）模型，通过调整等级损失函数来解决特定问题，因此获得了优秀的效果。Kang 等[26]分析了常用的两种用于顺序动态捕获的方法：马尔可夫链和循环神经网络。为兼顾前者简约的优势和后者在高密度数据集中表现优秀的特性，提出一种基于自我注意的顺序模型。该模型能够捕获长期语义，同时使用关注机制来进行短期预测。

综上所述，当前对于网络行为分析的研究主要应用于推荐系统并服务于移动式应用中，在电信网络诈骗犯罪防控领域应用较少。在网络行为分析的具体技术上，当前网络行为编码技术主要以挖掘网络流量的表面特征为主，存在特征冗余大且选取困难的问题；对用户行为倾向性的分析大多只考虑单次网络行为，缺少对内部特征的挖掘。因此，针对网络行为分析中特征间关系难以确定、特征选取困难、行为稀疏且行为间关系不确定、行为无法表达真实意图等问题，本文提出了 VIM-TFCN 模型。模型主要贡献：用随机森林算法结合行为特征关联，解决行为稀疏、关系不明确与某些行为无法表达真实意图的问题，达到精确地预测网络行为序列的被害性的目的。

3　问题定义

定义 1　（隐性行为特征）包括从网络流量的交互情况、数据量、时间 3 个角度提取的特征，用 feat 表示，feat∈Feature。

定义 2　（显性行为特征）指用户在进行单击、评论、注册、登录、交易等的行为特征序列，用 act 表示，act∈Feature。

定义 3　（网络行为编码）将每一条用户行为的特征序列进行行为编码，其中 feat 采用 *k*-means 算法进行离散化，不考虑连续数值的意义；act 则通过预先设立的字典，使用 label_binarize 二值化编码。重构特征序列 Feature 并表示为 action，action∈Action。以此得出行为矩阵 Action。

定义 4　（显隐性行为特征间的关联规则（rule））通过挖掘特征序列发现，当显隐性行为呈现出特定的组合后，该行为序列的被害性会大大增加。这种特定的组合关系被称作显隐性行为的关联规则。

定义 5　（关联规则特征序列）通过挖掘已知的被害行为序列显隐性行为，得到关联规则列表。被害与非被害行为序列全部使用关联规则列表进行重构，将重构后的特征序列称为关联规则特征序列。

问题定义　给定用户与服务器之间交互时所提取的行为特征矩阵，判断用户行为特征序列是否具有被害性。

4　模型设计

VIM-TFCN 模型主要包括两个部分：网络行为特征挖掘、电信网络诈骗被害性分析。模型流程如图 1 所示。

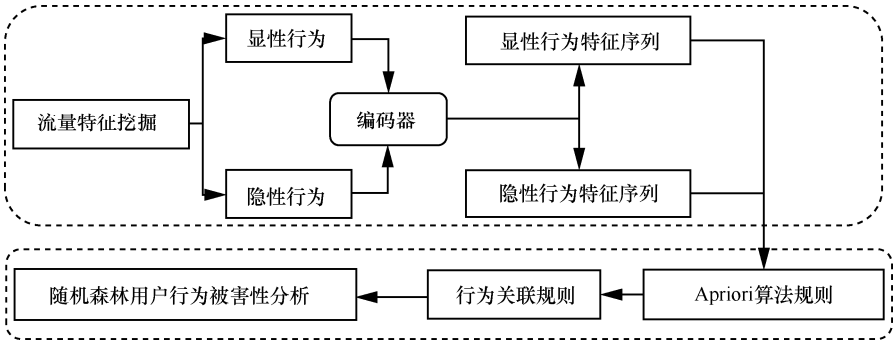

图 1　VIM-TFCN 模型流程

4.1　用户网络行为特征挖掘

用户在访问网站时存在显性与隐性的行为特征。通过显性与隐性网络行为特征研究，能够提高行为编码的准确性与模型的可靠性。

隐性行为特征序列主要从流量特征中提取，从时间、数量以及交互情况的角度分为 3 类。由于隐性特征无须挖掘其连续数值上的内容，因此将采集到的隐性特征数据进行离散化处理。

显性行为是抓取 HTTP 数据分组中的各字段内容，通过行为分类字典筛选获得，能够准确地反映用户在该网站进行的操作。本文显性行为特征仅考虑行为是否出现及其所属种类，并不考虑行为出现的频率与次数，因此采用 label_binarize 编码方式对显性行为特征进行编码。

初始行为特征参数见表 1。

表 1　初始行为特征参数

行为特征参数	描述
Srcip	源地址
Dstip	目标地址
final_byte_number	总字节
packet_number	总分组数
Dstport	目标端口
Keep_time	持续时长
send_packet	发送分组数
receive_packet	接收分组数
send_len	发送分组长度
receive_len	接收分组长度
sender_header_len	发送 tcp 头部信息长度
receiver_header_len	接收 tcp 头部信息长度
Load_size	负载总大小
Send_Load_size	发送负载大小
rec_Load_size	接收负载大小
Feature	行为特征

4.2　电信网络行为被害性分析算法

通过对实际数据的分析，可观测到用户访问诈骗网站与访问正常网站的网络行为具有一定的相似性，单一考虑用户单次网络行为的被害风险是片面的。因此，通过挖掘用户的显性行为特征与用户与服务器交互的隐性行为特征间的关系，能实现用户网络行为风险预测，显著增强网络行为同用户行为是否被害的关联性。显/隐性行为互相不存在直接映射关系，各自内部也不存在关联性；同时大部分显性特征行为同用户是否被害同样没有直接性的关联。但可以发现许多被害用户行为特征序列Feature中的若干特征指数明显较高的情况下发生了一些不指向电信网络诈骗的显性行为。

基于上述情况，本文提出一种网络行为被害性分析算法解决该问题。其中，行为关联规则挖掘示意图如图 2 所示。

步骤 1　获取行为频繁项，挖掘行为关联规则。初步的实验表明，被害行为序列同非被害行为序列的关联规则具有明显差异。在一个用户行

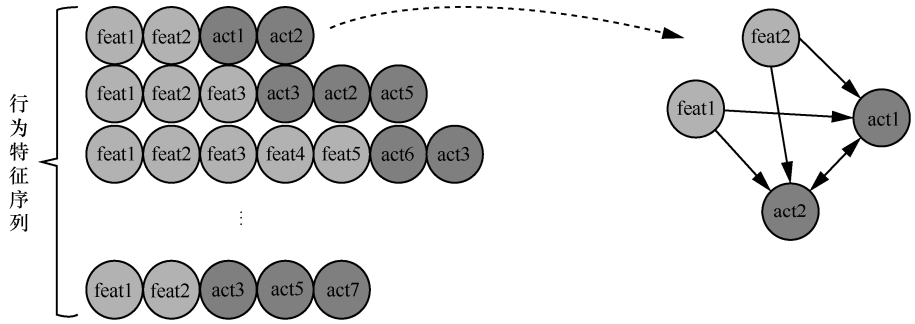

图 2　行为关联规则挖掘示意图

为序列中，单个 act 行为的出现无法表示该序列具有被害风险，但当该 act 行为与其他显/隐性特征关联关系大量满足从被害行为序列中提取的关联规则时，就可以在一定程度上证明该序列存在被害的风险。

获取频繁项。计算行为支持度，通过设定阈值生成行为频繁项。如针对图 2 中{feat1,feat2,act1}这一频繁项，支持度 S 计算式：

$$S = F(\text{feat}_1^i, \text{feat}_2^i) / N(\text{act}_1^i) \tag{1}$$

挖掘行为关联规则。{feat1,feat2}→ {act1}→ {true}中，{feat1,feat2→act2}作为频繁项具有关联性，据此计算置信度 C，将关联规则量化，设定关联规则置信度取值范围，达到排除低被害风险关联规则、提升规则有效性的目的，计算式如式（2）所示。

$$C = F(\text{feat}_1^{\text{i}}, \text{feat}_2^i \mid \text{act}_1^i) / F(\text{feat}_1^{\text{i}}, \text{feat}_2^i) \tag{2}$$

频繁项与关联规则的挖掘与量化表示，显著增强行为序列之间各项行为的关联性，其优点在于当行为序列出现某项无法表征用户实际意图的黑名单行为时，该序列不会被直接认为具有被害风险。同样，当行为序列中不存在黑名单行为时，也能判断其被害风险系数。

步骤 2　结合显隐性行为本身的特征与各行为特征间的内在关联规则，重构特征序列。原始行为特征序列用于直接表达存在的显/隐性行为特征，重构后的特征序列则用于表达该行为序列存在哪些可疑的关联关系。

步骤 3　采用随机森林对特征序列实现监督式机器学习，实现被害风险判断。重构后的特征序列由原始行为特征序列中存在的关联规则权重较大项组成，且已经标注是否为被害行为序列，这是监督学习的前提条件。算法使用自助法（bootstrap）对该特征序列组成的特征矩阵 $\boldsymbol{D}$ 进行有放回的随机抽样，生成包含 m 条特征序列的训练集，此过程重复 t 次，形成采样集 $D(t)$；然后针对每一个训练集构建决策树。随机森林算法示意图如图 3 所示。

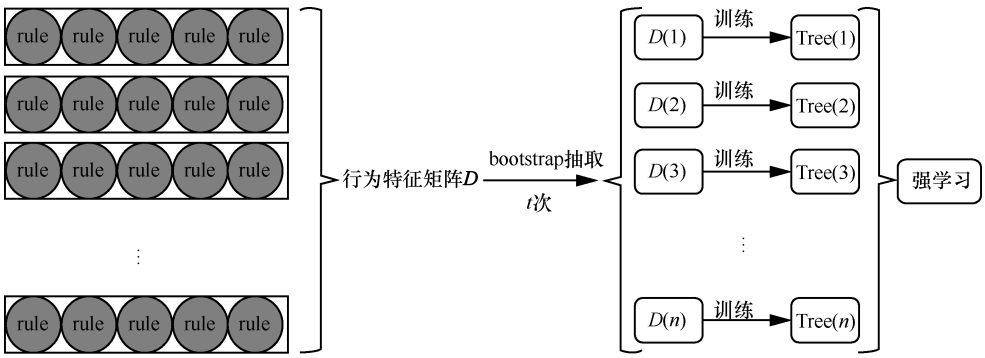

图 3　随机森林算法示意图

该算法的决策树在创建的过程当中，并不会拉取特征序列中所有的规则用于分裂，而是采取随机抽取的策略，从中获取最优解。通过合理设置每棵决策树拉取特征的数量与决策树的数量，可以在被害识别的准确度和模型运行效率间寻找最优的平衡点。

VIM-TFCN 模型通过挖掘网络行为关联规则，排除非意图表达行为，研究网络行为深层关系，判断行为序列被害性，从而达到增加被害性风险分析精确度的目的。

VIM-TFCN 算法伪代码如算法 1 所示。

算法 1 网络行为被害性分析算法 VIM-TFCN

输入 已标定行为序列数据集 D，待检测行为序列数据集 P

输出 网络行为被害性评估结果集合 Result

Begin

Result=[]//被害性评估结果集

TrainData=getTrainingData(*D*)//获取训练数据

TestData=getTestingData(*P*)//获取待检测数据

Rules=getAssociationRules(*D*)//生成关联规则列表

TrainSeq=getSeqbyRules(TrainData,Rules)//根据挖掘完毕的关联规则，获取训练行为序列集

TestSeq=getSeqbyRules(TestData,Rules)//根据关联规则，获取测行为序列集

Forest=TrainRandomForest(TrainSeq)//训练改进的监督式随机森林模型

For seq in TestSeq://对于每一条测试序列进行测试

Judgeresult=Forest.predict(seq)//通过随机森林模型获得风险评估结果

Result.append(Judgeresult)//将每一条风险评估结果放入被害性评估结果集

End

5 实验与分析

5.1 实验设置

本实验数据集主要包括被害人网络行为数据 2 051 条，非被害人网络行为数据 2 386 条。收集从公安部门获取的诈骗网站地址与公开的非诈骗网站地址，分类后模拟正常用户访问，使用 Wireshark 工具进行流量抓取，所得数据作为实验的数据集编写脚本从流量包中提取对每一个网站访问时产生的行为特征，形成原始特征序列。数据集网站种类与数量见表 2。

表 2 数据集网站种类与数量

诈骗类		非诈骗类	
种类	数量	种类	数量
博彩	534	网页游戏	453
虚假投资	730	金融投资	323
虚假论坛	364	生活交友	467
虚假借贷	507	彩票	156

选取以上类型的非诈骗网站，使诈骗网站与非诈骗网站所提取的网络行为序列具备相似性，从而验证本文结论。

本实验的实验环境设置如下。

数据库系统版本为 MySQL5，系统环境为内存 4 GB，处理器 8 个，操作系统为 kali_linux_2020.1，编程语言为 g++,python3.8.1。

本实验提取特征过程需要大量字典比对，且采用多线程模式提高效率，需要占用较大的内存，因此选取 4 GB 以上的运行内存；对 Wireshark 抓取的 pcapng 格式流量引用 Python 程序语言的 Pyshark 库批量化解析。

5.2 评价指标

（1）混淆矩阵

混淆矩阵（confusion matrix）是用来评价分类的标准方式，采用 $N\times N$ 的矩阵表示。在本文中，混淆矩阵为一个 2×2 矩阵，矩阵中每一列的总数

表示模型预测为该种类时数据的数目；每一行代表了数据的真实归属种类，每一行的数据总数表示该种类数据实例的数目。

（2）ROC 曲线

ROC 曲线指在特定条件下，以 FPR=*y*/*N* 为横坐标，表示负样本错误预测为正样本的比例；以 TPR=*y*/SN 为纵坐标，表示预测正确的正样本在所有正样本中所占比例。

对于 ROC 曲线，对角线称为纯机遇线，代表辨别力为 0；距离纯机遇线越远，辨别能力越好。

（3）精确率与召回率

用 TP 表示将正类预测为正类，FP 表示将负类预测为正类，FN 表示为正类预测为负类；精确率 *P*（Precision）计算式为：

$$P = \mathrm{TP} / (\mathrm{FP} + \mathrm{TP}) \tag{3}$$

召回率 *R*（Recall）计算式为：

$$R = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{4}$$

精确率与召回率通常呈现此消彼长的状况。可以通过调整提升度的阈值，确定最佳的关联规则表。

5.3　实验步骤与参数调优

5.3.1　实验步骤

（1）对原始特征序列中的行为特征字段采用 label_binarize 二值化编码，该字段由流量中的各项具有行为代表性的单词按照发生先后顺序组成；其中行为字典构建是基于对诈骗网站各类变量的常用命名方式、相关英文与中文拼音的组合。行为特征字段中的各单词在去重、统一小写、去除特殊字符等操作后采用最大公共串与行为字典进行匹配识别显性行为种类，并加入行为特征序列。其中，设定 *L* 大于 0.5。*L* 越大表示显性行为更精确，但会出现行为无法识别的现象。*L* 计算式如式（5）所示。

$$L = \text{最大字串长度/行为单词} \tag{5}$$

（2）对表 1 中的隐性行为特征序列进行离散化。实验采用 *K*-means 算法对隐性行为特征进行离散化处理，设为 5 个等级。对于每一项隐性行为特征使用 0~5 的数值表示，删去原有的具体值。

（3）采用 Apriori 算法对显/隐性行为特征序列进行关联规则挖掘，设定提升度为大于 0.9 减小开销，最终挖掘规则共计 600 余条。提升度越大，规则越多，精度越高，但开销显著增加。关联规则参数见表 3。

表 3　关联规则参数

关联规则参数	描述
Antecedents	规则前项
Consequents	规则后项
antecedent support	前项支持度
consequent support	后向支持度
Support	前项对后项支持度
Confidence	置信度
Lift	提升度
Leverage	关联性参数 1
Conviction	关联性参数 2

（4）通过挖掘的关联规则来对于每一条行为序列进行特征重构。先使用关联规则的编号与权重表达每一条行为序列的关联规则搭配规律；然后使用随机森林算法对重构后特征进行机器学习，识别网络行为被害性。

（5）本文所提方法 VIM-TFCN 与另外两种分类算法进行对比，可在 Github 获取开源代码。

- C45 算法：ID3 算法的扩展，通过决策树找到行为特征与属性的映射，对未知个体分类进行识别。
- 贝叶斯算法：对缺失数据不敏感，假设行为特征相互独立，结合先验概率与后验概率进行分类，避免过度拟合。

5.3.2　参数调优

在随机森林算法的实际应用当中，子决策树的数量（Numtree）与决策树拉取特征数

（Numfeature）对分类效果影响较大。随着子决策树数量和拉取特征数变多，分类的精度会呈现先上升后下降,最后趋于稳定的情况，且模型运行速度与性能会显著降低。根据控制变量的原则，通过调整参数大小来对分类结果进行对比。控制变量参数调优结果见表 4。

随着 Numfeature 参数增大，精准率与召回率上下浮动，MCC 区域呈现减小的趋势，在同等结果下选取内存开销较小值，确定该参数值为 4；随着 Numfeature 参数增大，精准率、召回率、MCC 区域呈现波动上升，最后趋于稳定，内存开销与时耗呈现指数增加。根据实验结果确定 Numfeature 参数为 4，Numtree 参数为 50~70 时达到最优。

5.4 实验结果分析

VIM-TFCN 模型同朴素贝叶斯、C45 分类算法在用户网络行为被害性识别效果对比见表 5 和图 4 所示。

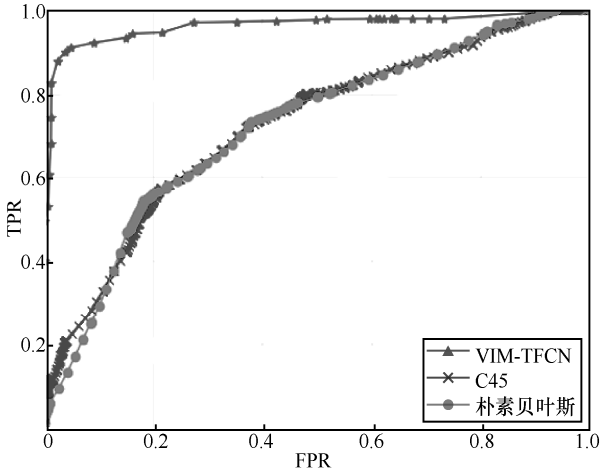

图 4 ROC 曲线对比

由图 4 可知，VIM-TFCN 模型的 ROC 曲线贴近 TPR 轴，而 C45 模型和朴素贝叶斯模型的 ROC

表 4 控制变量参数调优结果

决策树参数		分类效果变量					
Numfeature	Numtree	TP	FP	Precision	Recall	F-Measure	MCC
0	100	0.964	0.095	0.965	0.964	0.964	0.959
2	100	0.963	0.096	0.964	0.963	0.963	0.958
4	100	0.964	0.095	0.965	0.964	0.964	0.959
6	100	0.963	0.096	0.964	0.963	0.963	0.956
8	100	0.963	0.096	0.964	0.963	0.963	0.956
4	20	0.962	0.097	0.963	0.962	0.962	0.955
4	40	0.964	0.095	0.965	0.964	0.964	0.959
4	60	0.965	0.094	0.966	0.965	0.965	0.960
4	80	0.964	0.095	0.965	0.964	0.964	0.959
4	100	0.964	0.095	0.965	0.964	0.964	0.959

表 5 混淆矩阵对比

Confusion Matrix		VIM-TFCN 预测值		C45 模型预测值		朴素贝叶斯预测值	
		TRUE	FALSE	TRUE	FALSE	TRUE	FALSE
真实值	TRUE	2 005	46	1 648	403	1 372	679
	FALSE	32	2 254	443	1 843	359	1 927

曲线更贴近对角线。由表 5 混淆矩阵可得，VIM-TFCN 模型的精确率为 0.982，召回率为 0.984；C45 模型和朴素贝叶斯模型的精确率分别为 0.804 和 0.760，召回率分别为 0.788 和 0.798。根据对 ROC 曲线以及混淆矩阵的评估分析可知，通过显/隐性行为的关联规则组合判断连续行为被害性的效果，优于通过一项或多项行为判断连续行为被害性的效果。

6　结束语

本文针对电信网络诈骗犯罪中网络行为特征选取表面化、特征间内在规则挖掘不足、网络行为稀疏、行为序列间关系难以确定等问题，提出基于网络行为流量分析的电信网络诈骗犯罪被害性识别模型，分析各类网络行为的内在相关性，深度挖掘显/隐性网络行为，达到识别网络行为是否具有被害性的目的。通过公安机关被害人网络行为数据分析验证，模型可以有效地进行被害人网络行为分类，识别网络行为是否具有被害性。下一步研究将加入特征存取栈模块，实现实时监控异常网络行为，并增强模型鲁棒性，发掘更深层次的关联规则。

参考文献：

[1] 佟晖, 唐卫中, 蔡家艳, 等. 电信诈骗态势与反诈新思路研究[J]. 北京警察学院学报, 2021(1): 1-14.
TONG H, TANG W Z, CAI J Y, et al. Research on the situation of telecom fraud and new ideas of anti fraud[J]. Journal of Beijing Police College, 2021(1): 1-14.

[2] 周坚, 石永革, 何美斌. 基于 A-D 模型的 *K*-means 算法在通话异常客户挖掘中的应用[J]. 电信科学, 2018, 34(4): 81-89.
ZHOU J, SHI Y G, HE M B. Application of *K*-means algorithm based on A-D model in calling abnormal customer mining[J]. Telecommunications Science, 2018, 34(4): 81-89.

[3] 李力卡, 马泽雄, 陈庆年, 等. 电话诈骗防治技术解决方案与运维对策研究[J]. 电信科学, 2014, 30(11): 166-172.
LI L K, MA Z X, CHEN Q N, et al. Research of technology solutions and operation countermeasures to telephone fraud prevention and control[J]. Telecommunications Science, 2014, 30(11): 166-172.

[4] 王海坤, 潘嘉, 刘聪. 语音识别技术的研究进展与展望[J]. 电信科学, 2018, 34(2): 1-11.
WANG H K, PAN J, LIU C. Research development and forecast of automatic speech recognition technologies[J]. Telecommunications Science, 2018, 34(2): 1-11.

[5] 张蕾, 张鹏, 孙伟, 等. 面向高速网络流量的恶意镜像网站识别方法[J]. 通信学报, 2019, 40(7): 87-94.
ZHANG L, ZHANG P, SUN W, et al. IMM4HT: an identification method of malicious mirror website for high-speed network traffic[J]. Journal on Communications, 2019, 40(7): 87-94.

[6] 韩浩, 刘博文, 林果园. 基于改进的 TrustRank 算法的钓鱼网站检测[J]. 电信科学, 2018, 34(3): 86-94.
HAN H, LIU B W, LIN G Y. Detection of phishing websites based on the improved TrustRank algorithm[J]. Telecommunications Science, 2018, 34(3): 86-94.

[7] 臧小东, 龚俭 ,胡晓艳. 基于 AGD 的恶意域名检测[J]. 通信学报, 2018, 39(7): 15-25.
ZANG X D, GONG J, HU X Y. Detecting malicious domain names based on AGD[J]. Journal on Communications, 2018, 39(7): 15-25.

[8] 韩春雨, 张永铮, 张玉. Fast-flucos: 基于 DNS 流量的 Fast-flux 恶意域名检测方法[J]. 通信学报, 2020, 41(5): 37-47.
HAN C Y, ZHANG Y Z, ZHANG Y. Fast-flucos: malicious domain name detection method for Fast-flux based on DNS traffic[J]. Journal on Communications, 2020, 41(5): 37-47.

[9] ZHOU S L , WANG X , YANG Z R . Monitoring and early warning of new cyber-telecom crime platform based on BERT migration learning[J]. China Communications, 2020, 17(3): 140-148.

[10] ZOLOTUKHIN M, HÄMÄLÄINEN T, KOKKONEN T, et al. Analysis of http requests for anomaly detection of Web attacks[C]//Proceedings of 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing. Piscataway: IEEE Press, 2014: 406-411.

[11] YU Y, LIU G, YAN H, et al. Attention-based Bi-LSTM model for anomalous HTTP traffic detection[C]//Proceedings of 2018 15th International Conference on Service Systems and Service Management. Piscataway: IEEE Press, 2018: 1-6.

[12] YANG W, ZUO W, CUI B. Detecting malicious URLS via a keyword-based convolutional gated-recurrent-unit neural network[J]. IEEE Access, 2019(7): 29891-29900.

[13] PARK S, KIM M, LEE S. Anomaly detection for HTTP using convolutional autoencoders[J]. IEEE Access, 2018(6): 70884-70901.

[14] ARZHAKOV A V, TROITSKIY S S, VASILYEV N P, et al. Development and implementation a method of detecting an attacker with use of HTTP network protocol[C]//Proceedings of 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering. Piscataway: IEEE Press, 2017: 100-104.

[15] THANG T M, KIM J. The anomaly detection by using DBSCAN clustering with multiple parameters[C]//Proceedings of 2011 International Conference on Information Science and Applications. Piscataway: IEEE Press, 2011: 1-5

[16] ZHANG M, LU S, XU B. An anomaly detection method based on

multi-models to detect Web attacks[C]//Proceedings of 2017 10th International Symposium on Computational Intelligence and Design. Piscataway: IEEE Press, 2017(2): 404-409.

[17] ERFANI S M, RAJASEGARAR S, KARUNASEKERA S, et al. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning[J]. Pattern Recognition, 2016(58): 121-134.

[18] 连一峰, 戴英侠, 王航. 基于模式挖掘的用户行为异常检测[J]. 计算机学报, 2002(3): 325-330.
LIAN Y F, DAI Y X, WANG H. Anomaly detection of user behaviors based on profile mining[J]. Chinese Journal of Computers, 2002(3): 325-330.

[19] 田新广, 孙春来, 段洣毅, 等. 基于机器学习的用户行为异常检测模型[J]. 计算机工程与应用, 2006(19): 101-103, 111.
TIAN X G, SUN C L, DUAN M Y, et al. Model of anomaly detection of users behaviors based on machine learning[J]. Computer Engineering and Applications, 2006(19): 101-103, 111.

[20] 陈胜, 朱国胜, 祁小云, 等. 基于深度神经网络的自定义用户异常行为检测[J]. 计算机科学, 2019, 46(S2): 442-445, 472.
CHEN S, ZHU G S, QI X Y, et al. Custom user anomaly behavior detection based on deep neural network[J]. Computer Science, 2019, 46(S2): 442-445, 472.

[21] 胡富增, 王勇军. 基于数据挖掘的计算机用户行为分析与识别[J]. 自动化技术与应用, 2020, 39(6): 42-47.
HU F Z, WANG Y J. Analysis and recognition of computer user behavior based on data mining[J]. Techniques of Automation and Applications, 2020, 39(6): 42-47.

[22] WANG S, CAO L, WANG Y. A survey on session-based recommender systems[J]. arXiv: 1902. 04864, 2019.

[23] TANG J, WANG K. Personalized top-n sequential recommendation via convolutional sequence embedding[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2018: 565-573.

[24] SUN F, LIU J, WU J, et al. BERT4Rec: sequential recommendation with bidirectional encoder representations from transformer[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2019: 1441-1450.

[25] HIDASI B, KARATZGLOU A, BALTRNAS L, et al. Session-based recommendations with recurrent neural networks[J]. arXiv: 1511. 06939, 2015.

[26] KANG W C , MCAULEY J . Self-attentive sequential recommendation[C]//Proceedings of 2018 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2018.

[作者简介]

周胜利（1982− ），男，博士，浙江警察学院硕士生导师，主要研究方向为大数据安全、机器学习。

徐啸炀（1999− ），男，浙江警察学院在读，主要研究方向为网络安全与机器学习。