

基于时空行为与静态属性分析的陌陌应用恶意用户检测方法

杨海峰^{1,2}, 杜少勇², 魏果², 时文旗², 罗向阳²

(1. 郑州大学网络空间安全学院, 河南 郑州 450002; 2. 河南省网络空间态势感知重点实验室, 河南 郑州 450001)

摘要: 陌陌是一款流行的移动社交应用, 其中存在实施诈骗、网赌网贷等网络违法犯罪活动的恶意用户, 发现陌陌中的恶意用户具有重要意义。然而, 陌陌应用具有用户间社交关系松散、发布文本内容有限等特点, 导致现有恶意用户检测方法难以准确刻画用户特征, 检测准确率低。为此, 提出一种基于用户时空行为与静态属性分析的恶意用户检测方法。基于陌陌提供的位置服务, 搜集用户在线时间、位置、异常标签等数据, 构建用户时空信息数据集; 对比数据集中正常用户与具有异常标签的恶意用户的在线时间和位置分布差异, 分析恶意用户的时空行为特征; 基于用户在不同时间段的在线概率刻画用户的时间行为特征, 基于用户历史位置分布的聚集程度刻画用户的空间行为特征; 最后, 融合已获得的时空行为特征与性别属性训练用户分类模型, 检测恶意用户。基于收集的67 280个陌陌用户的1 894 917条数据, 开展了恶意用户检测实验, 结果表明, 提出的方法能够有效检测出恶意用户, 特别是针对历史发布内容较少的恶意用户时, 所提方法正确分类恶意用户的能力显著优于代表性方法SybilSCAR和DatingSec, AUC-ROC和AUC-PR分别提升4.6%以上和19.88%以上。

关键词: 陌陌; 恶意用户检测; 时空行为; 机器学习

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024060

Malicious user detection in Momo application based on spatio-temporal behavior and static attribute analysis

YANG Haifeng^{1,2}, DU Shaoyong², WEI Guo², SHI Wenqi², LUO Xiangyang²

1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China

2. Key Laboratory of Cyberspace Situation Awareness of Henan Province, Zhengzhou 450001, China

收稿日期: 2024-01-30; 修回日期: 2024-07-28

通信作者: 时文旗, shiwenqi1606@163.com

基金项目: 国家重点研发计划(2022YFB3102900); 国家自然科学基金(U23A20305, 62172435, 62372465, 62002386); 河南省重点研发项目(221111321200); 河南省自然科学基金(232300421098)

Foundation Items: The National Key Research and Development Program of China (2022YFB3102900), The National Natural Science Foundation of China (U23A20305, 62172435, 62372465, 62002386), Key Research and Development Project of Henan Province (221111321200), The Henan Province Science Foundation for Youths (232300421098)

引用格式: 杨海峰, 杜少勇, 魏果, 等. 基于时空行为与静态属性分析的陌陌应用恶意用户检测方法[J]. 网络与信息安全学报, 2024, 10(4): 143-158.

Citation Format: YANG H F, DU S Y, WEI G, et al. Malicious user detection in Momo application based on spatio-temporal behavior and static attribute analysis[J]. Chinese Journal of Network and Information Security, 2024, 10(4): 143-158.

Abstract: Momo, a popular mobile social application, has attracted a significant number of malicious users engaged in illegal online activities such as fraud, online gambling, and illegal lending. The detection of these malicious users is of great importance. However, the loose social relationships between users and the limited textual content on Momo have made it challenging for existing methods to accurately characterize user features, resulting in low detection accuracy. To address this issue, a malicious user detection method based on the analysis of user spatio-temporal behavior and static attributes was proposed. Firstly, a user spatio-temporal information dataset was constructed by collecting user online time, location, and abnormal labels data provided by Momo's location service. Then, the differences in online time and location distribution between normal users and malicious users with abnormal labels in the dataset were compared to analyze the spatio-temporal behavior characteristics of malicious users. Subsequently, the temporal behavior characteristics of users were characterized by the online probability of users in different time periods, and the spatial behavior characteristics of users were characterized by the degree of clustering of users' historical location distribution. Finally, the obtained spatio-temporal behavior features and gender attributes were combined to train a user classification model for malicious user detection. A malicious user detection experiment was conducted based on 1,894,917 data points collected from 67,280 Momo users. The results demonstrate that the proposed method could effectively detect malicious users, especially for those with fewer historical content publications. The proposed method exhibits significantly better performance in correctly classifying malicious users compared to representative methods such as SybilSCAR and DatingSec, with an improvement of over 4.6% in AUC-ROC and over 19.88% in AUC-PR.

Keywords: Momo, malicious user detection, spatio-temporal behavior, machine learning

0 引言

移动互联网的快速发展促进了移动社交应用的广泛流行,陌陌、微信等移动社交应用拥有大量的用户^[1-3]。为了满足用户交友需求,陌陌应用会提供基于位置的用户发现(LBSD, location-based social discovery)服务^[4-6]。该服务允许用户发现自身所处地理位置附近的其他用户,并能够查看附近用户的公开资料、相对距离等信息。然而,LBSD服务在方便用户交友的同时,也为恶意用户提供了犯罪渠道,不法分子可以利用LBSD服务发现附近用户,并通过即时消息进行诈骗、散布谣言^[7]。此外,LBSD服务展示附近用户相对距离的特性,也使用户存在真实位置被恶意用户窃取的风险^[8-9]。因此,如何高效准确地检测出移动社交应用中的恶意用户,对保护用户隐私及安全、预防网络犯罪等具有重要的现实意义。

当前,社交平台中的恶意用户通常被定义为骚扰他人或危害他人利益的用户^[10]。在陌陌应用中,部分用户主页带有“异常”标签,该标签由其他用户举报或陌陌内置的异常检测机制所标记并显示,这为开展恶意用户检测模型的训练和检测性能的评估提供了参考。

根据利用的数据类型不同,现有的在线社交平台恶意用户检测方法分为两类。第一类方法是通过构建用户间的社交关系图,基于恶意用户与正常用户间的社交结构差异识别恶意用户^[10-15]。第二类方法是基于恶意用户与正常用户在发布内容或行为方面的差异,针对性地提取用户属性或行为特征,训练恶意用户分类模型,从而检测恶意用户^[16-22]。然而,陌陌应用作为陌生人交友平台,用户间社交关系呈现随机性,极为松散,这导致基于社交关系图的检测方法难以适用;此外,当恶意用户发布内容较少或不发布内容时,现有基于用户发布内容和行为的检测方法难以检测出此类恶意用户,即使恶意用户正在利用陌陌应用的即时通信功能进行造谣、诈骗等违法行为。

针对上述问题,本文利用陌陌LBSD服务,搜集了一批用户的公开资料数据,并对所搜集数据中用户的时空行为进行分析,发现带有“异常”标签的恶意用户具有差异化的在线时间偏好与位置分布特征,且性别为女性的账号在恶意用户中占比高达75.9%。基于该发现,本文提出了一种基于用户时空行为信息的陌陌应用恶意用户发现方法。该方法通过计算用户在不同时间段的

在线概率来描述用户的在线时间偏好, 作为用户的时间行为特征。同时, 使用聚类算法计算用户的历史位置坐标形成的簇数量, 并统计平均簇内坐标间距离来描述用户历史位置分布的聚集状态, 作为用户的空间行为特征。最后, 聚合用户的时空行为特征与性别属性, 选取合适的分类算法, 训练恶意用户分类模型。实验结果表明, 所提方法能够有效检测出潜在的恶意用户, 且当用户历史发布文本内容较少时, 所提方法明显优于现有代表性方法。

本文主要贡献如下。

1) 发现了恶意用户的差异化时空行为特征。本文设计并实现了一套用户公开资料自动获取程序, 在多个城市搜集了 67 280 个陌陌用户的公开资料数据。通过对比分析正常用户与恶意用户的在线时间偏好及坐标分布特征, 揭露了恶意用户的时空行为规律, 并以此构建恶意用户的时空行为特征。

2) 设计了一种融合用户时空行为特征和性别属性的恶意用户检测方法。该方法利用陌陌用户广泛具有的在线时间和位置信息来表征用户特征, 可以有效降低稀疏的用户间社交关系和有限的历史发布文本内容对检测性能的影响。实验结果表明, 所提方法在恶意用户发布的历史动态数据有限的情况下, 明显优于具有代表性的方法。

1 相关工作

1.1 社交平台恶意用户检测

在检测在线社交平台恶意用户的历史工作中, 主要包括基于社交关系图和基于用户发布内容或行为两类方法。

第一类方法是利用用户间社交关系构建图结构, 其以恶意用户与正常用户间联系结构不同为前提来判断恶意用户。Cao 等^[10]将随机游走策略引入社交关系图, 用于在大型在线社交网络中发现恶意用户。Gong 等^[11]提出了 SybilBelief, 这是一种半监督学习框架, 通过无监督学习算法对社交网络中的节点进行聚类分析, 提取多种特征训练半监督学习模型, 用来区分正常节点和恶意节点。SybilSCAR^[12]和 SYBILFUSE^[13]结合了基于随机游走和循环置信传播方法的优点, 以提高恶意用户分类准确率。Li 等^[14]利用社交关系网

络的异构性, 通过分析不同类型节点和边之间的关系来识别恶意用户, 该方法适用于不同规模的社交网络。Wang 等^[15]提出了一个协作分类框架, 利用网络结构信息推断未知节点的标签, 并结合联合权重学习和节点传播特性来检测恶意用户。

第二类方法是基于用户发布内容或行为, 通过恶意用户特定的生成内容或行为特征来检测恶意用户。Al-Qurishi 等^[16]通过分析用户发布内容、社交关系和用户基础资料 3 个方面的特征, 来检测大规模社交网络中明显偏离群体行为的恶意用户。He 等^[17]将用户发布内容和注意力机制应用于恶意用户检测工作中, 通过挖掘文本内容中潜在的信息提升恶意用户检测性能。Shen 等^[18]根据用户基础资料与行为交互特征建立了一个信任感知框架, 能够有效阻止恶意用户的早期不法活动。Breuer 等^[19]发现新用户最初的好友请求行为与用户是否异常具有一定关联, 通过新账户发起或接收的好友请求来判断其是否恶意用户。Xu 等^[20]通过聚合社交关系图中直接用户和间接用户的特征, 采用多阶段多任务学习范式, 实现多种类型的恶意账号高精度分类。Saeed 等^[21]系统性地针对 Reddit 应用中恶意煽动舆论的账户进行分析, 通过内容与行为相似性发现潜在的恶意用户。Zhu 等^[22]针对网络借贷服务中的欺诈账户, 通过发掘大量低密度用户信息中的有价值信息提高信息密度, 以提升欺诈账户检测能力。

1.2 基于 LBSD 服务的用户定位

移动社交应用提供的 LBSD 服务已被大量工作证实存在用户地理位置隐私泄露问题, 同时这些工作也在不断地促进相应平台增强位置隐私保护策略。Ding 等^[8]针对报告距离 (LBSD 服务返回的附近用户距离) 存在随机误差问题, 通过大量报告距离与真实距离的映射数据, 确定不同报告距离对应的真实距离边界, 然后采用三边测量方法, 实现对目标用户的粗粒度定位。为了实现精确定位, Wang 等^[9]将定位过程分为逼近和定位两个环节。逼近环节, 首先部署多个观测点, 然后选取最优观测点作为初始点; 定位环节, 利用初始点对目标的投影进行定位。该方法在仿真环境下的定位精度较高,

但在实际环境下存在开销大和定位成功率低的问题。Li 等^[23]提出基于空间划分的定位方法,通过对最小报告距离对应地理区间的二分查找,实现对实际环境中微信用户的高精度定位。在此基础上,文献[24-25]通过增加 LBSD 服务请求次数进一步提高了定位精度。针对报告距离随机化问题, Peng 等^[26]设计了一种基于数论的定位方法,提高了在实际环境下的用户定位精度。Polakis 等^[27]对用户定位问题进行了模型化描述,提出一个用户定位框架,并成功实现了对 Facebook Messenger、Skout 等多款移动社交应用用户的定位。Shi 等^[28]针对微信应用提出了一种基于加权最小二乘的定位方法,该方法通过对报告距离与真实距离的统计分析来确定报告距离与真实距离边界映射关系,然后使用最小二乘法与真实距离边界加权策略计算目标的理论位置,在保证定位精度的同时优化了定位效率。

2 背景介绍

本节介绍利用陌陌 LBSD 服务计算用户位置坐标的方法,同时对本文要解决的问题进行定义。

2.1 利用 LBSD 服务计算陌陌用户位置坐标

在本文中,运行着陌陌应用与数据搜集程序的安卓模拟器称为探针,并且可通过虚拟位置的方法将探针位置模拟至任意位置坐标,以搜集该位置附近的陌陌用户信息。如图 1 所示,将探针部署在 3 个不同的坐标点,控制探针向陌陌应用的 LBSD 服务器请求目标用户距离 d_1 、 d_2 、 d_3 ,分别以 3 个探针的坐标为圆心,以 3 个探针分别获取的距目标用户的距离为半径形成 3 个圆,将 3 个圆交汇处所形成的不规则区域的重心作为目标用户的位置坐标计算结果。

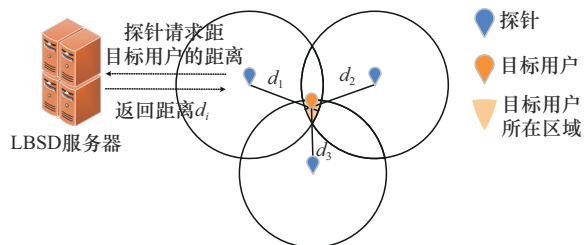


图 1 利用 LBSD 服务构建三边测量模型计算用户坐标

Figure 1 Constructing a trilateral measurement model using LBSD service to calculate user coordinates

2.2 问题定义

给定用户 u_i , 其属性由 $\{\delta, \varepsilon, x\}$ 描述, 根据属性集合输出 u_i 是否为恶意用户的预测值 $\hat{y}_i$, $\hat{y}_i = 1$ 表示 u_i 为恶意用户, $\hat{y}_i = 0$ 表示 u_i 为正常用户, 且 $\hat{y}_i \in \{0, 1\}$ 。本文常用符号及其含义如表 1 所示。

表 1 符号及其定义

Table 1 Symbols and definition

符号	定义
$\mathcal{V}$	用户集合
δ	探针距用户的距离集合
ε	用户的在线时间集合
x	用户性别
u_i	用户 i
p_i	探针 i
L	在线概率数组
ANN	平均最近邻指数
C	坐标集合
T	待定位记录集合
T'	仅包含两条记录的待定位记录集合
η	坐标构成的簇个数
γ	簇内坐标最小生成树距离
y_i	用户是否为恶意用户的标签
$\hat{y}_i$	用户是否为恶意用户的预测值

3 数据集构建

本节对陌陌用户信息搜集方法及本文所依托的数据集进行介绍。

3.1 陌陌用户信息搜集

本文使用安卓模拟器与 Appium 框架实现数据搜集工作, 安卓模拟器作为陌陌应用程序的运行环境, 使用基于 Appium 框架编写的数据搜集程序对页面上的用户信息进行提取。此外, 为了获取用户的在线时间, 将每个探针设置为只查看当前在线用户。

如图 2 所示, 对于陌陌应用的“附近”页面, 整个页面内容可以看作一个树状结构, 页面上的文本被封装于树结构的节点中, 可通过节点的唯一标识符 (节点 ID) 或节点路径 (xpath 值) 定位节点位置, 从而获取节点的内容信息。页面上的任意文本信息均可通过此方式获取。

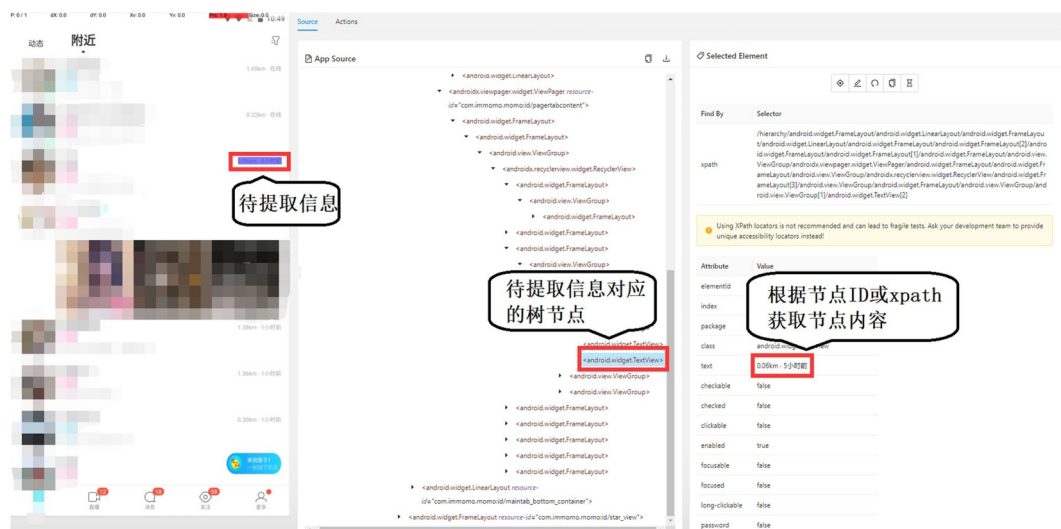


图2 使用Appium提取陌陌应用的页面信息

Figure 2 Extract Momo App page information with Appium

3.2 数据集介绍

基于上述方法, 本文通过虚拟探针位置, 针对郑州、上海和成都等的部分区域进行了长期的用户信息搜集工作, 形成数据集 $D_{ZhengZhou}$ 、 $D_{ShangHai}$ 和 $D_{ChengDu}$ 。搜集工作的时间段分别为2022年6月14日至2022年8月12日、2023年8月28日至2023年9月26日, 累计搜集了67 280个陌陌用户的1 894 917条数据, 数据的内容及格式如表2所示。另外, 本文对用户的发帖、评论及评论者相关信息等陌陌用户公开信息进行了记录, 由于这些用户信息数据与本文提出的方法无关, 因此不计入上述的数据统计条数, 这些数据可为之后的研究服务。最后, 为了保护用户隐私, 不详细说明具体地点, 并且所有数据均保存在专用离线计算机上。最重要的是, 本文将能够唯一标识用户身份的用户ID进行了匿名化处理, 使数据搜集工作不会侵犯到任何用户隐私。

表2 数据的内容及格式
Table 2 Data content and format

探针ID	用户ID	距离/m	性别	...	在线时间/s
p_1	u_1	1 100	男	...	1 657 111 242
p_2	u_2	900	女	...	1 657 111 264
...	...	...	...	...	...
p_n	u_n	1 000	女	...	1 657 112 276

4 用户时空行为分析

本节对数据集 $D_{ZhengZhou}$ 中正常用户与带有异常标签的恶意用户的在线时间和位置坐标分布特征进行对比分析。

4.1 用户时间行为分析

为了直观地观察恶意用户的在线规律, 图3分别展示了4天中每小时正常/恶意用户人数占正常/恶意用户总数的比例。可以看到恶意用户在05:00~12:00时间段的在线比例明显低于正常用户, 而在00:00~05:00与12:00~24:00时间段的在线比例显著高于正常用户。这是因为恶意用户将登录陌陌应用视为一种工作, 其在线活动时间遵循着相对固定的规律, 尤其是在凌晨时刻更为活跃。因此, 本文考虑利用恶意用户特定的在线时间偏好识别恶意用户。为此, 本文将通过描述用户的在线时间偏好来刻画用户的时间行为特征。

4.2 用户空间行为分析

本节首先利用三边测量方法计算用户的历史位置坐标, 然后对正常用户与恶意用户的历史位置分布模式进行对比分析。

4.2.1 坐标计算

基于本文数据集中探针距用户的距离信息可以计算用户的历史位置坐标。给定用户 u_i 所属的记录集合 $\delta = \{(t_1, d_{p1}), (t_2, d_{p2}), \dots, (t_n, d_{pi})\}$ 。其中, (t_1, d_{pi}) 表示 t_1 时刻探针 p_i 距用户 u_i 的距离为 d_{pi} 。

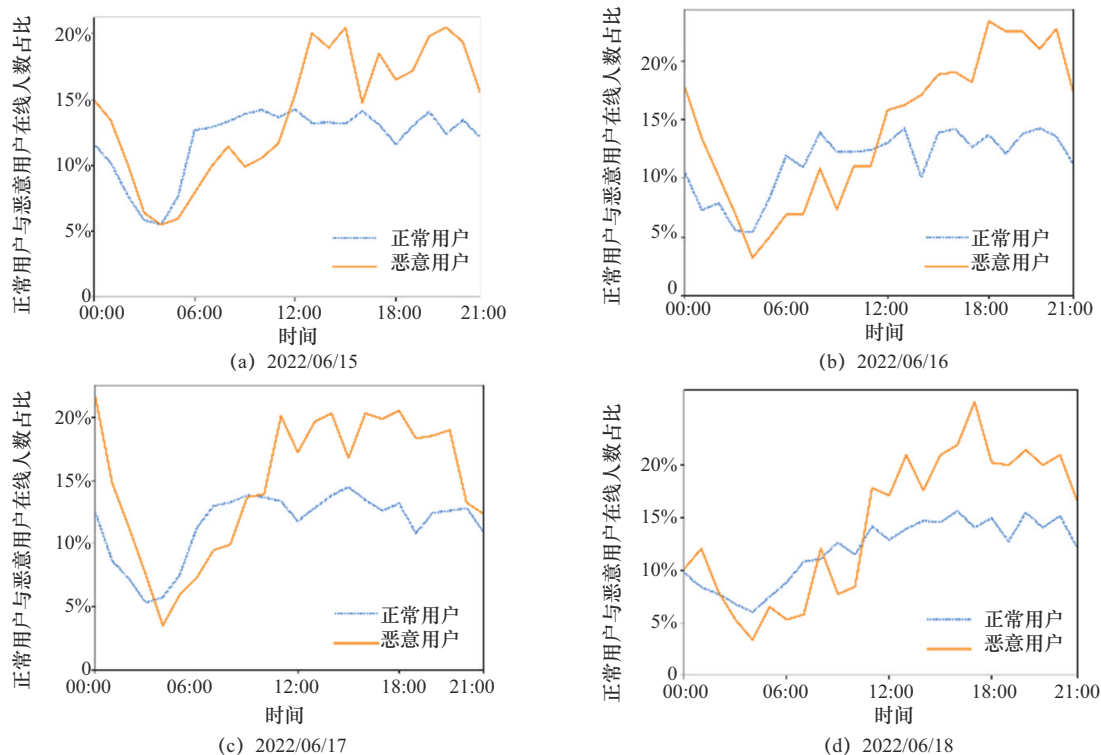


图3 4天中不同时间段内正常用户与恶意用户在线人数占比

Figure 3 The proportion of online activities between normal and malicious users during different time periods within 4 days

首先遍历 δ , 找出时间上相邻且所属3个不同探针的记录组合 $T = \{(t_1, d_{p_1}), (t_2, d_{p_2}), (t_3, d_{p_3})\}$; 然后, 如图4所示, 使用三边测量方法计算用户坐标, 并取以探针与目标用户的距离为半径所形成3个圆的相交区域重心作为定位结果。大量实测结果表明, 陌陌用户的定位误差在60 m以内, 这对分析用户的历史位置分布特征无实际影响。该方法得出的用户往往存在多组 T , 使用上述策略可以计算出 u_i 的历史位置坐标集合。

表3给出了一种在计算坐标过程中可能会发生的情况, 可以看到 p_1 与 p_3 分别于08:19和10:19

发现 u_1 , 在这2h内无法保证 u_1 始终处于静止状态, 这会产生错误的定位结果。为了减小这种情况对定位结果的影响, 需要保证 T 中任意两条记录时间差小于1h。

表3 探针在不同时间探测到同一用户
Table 3 The probes detect the same user at different times

探针ID	用户ID	在线时间	距离/m
p_1	u_1	2022/08.01 08:19	800
p_2	u_1	2022/08.01 08:40	1 100
p_3	u_1	2022/08.01 10:19	950

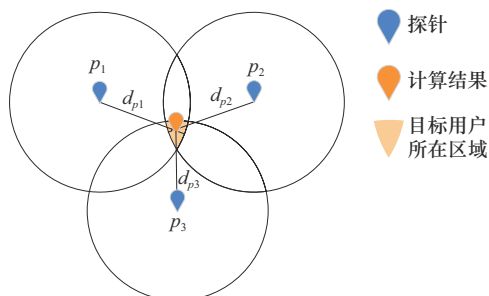


图4 使用三边测量方法计算坐标

Figure 4 Calculate coordinates using the trilateral measurement method

4.2.2 坐标分布分析

本节首先通过绘制正常用户与恶意用户的历史位置坐标点密度图, 分析正常用户与恶意用户的历史位置分布差异。考虑恶意用户数量较少, 为了更准确地比较正常用户与恶意用户的历史位置分布模式, 需平衡双方的用户数量与历史位置坐标数量。具体来说, 首先, 根据历史位置坐标数量对恶意用户群体进行分组, 每组间坐标数量相差5; 接着, 对于每一组, 根据该组内的恶意

用户数量及该组对应的坐标数量区间，在正常用户群体中选取相同数量的正常用户。这种做法可以使同一组内正常用户与恶意用户的数量与历史位置坐标数量保持相对平衡。通过上述方法，本文共获得恶意用户的 79 422 个历史位置坐标和正常用户的 79 714 个历史位置坐标。最后，基于获得的坐标绘制点密度图，如图 5 所示，颜色越深表示坐标点越密集。可以看到在图 5 的地理空间中，恶意用户的历史位置坐标总是聚集在正常用户的历史位置坐标聚集区域附近，且密度更高，呈现出了不同的分布模式。

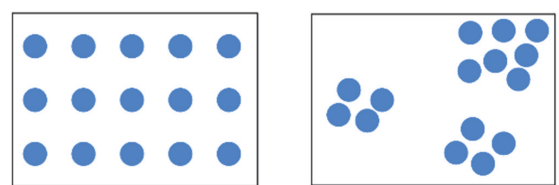
基于上文得出的恶意用户的历史位置坐标更加聚集的初步结论，本文使用平均最近邻指数^[29](ANN)来进一步描述用户位置坐标聚集程度。具体计算方式如下：

$$ANN = \frac{\dot{D}_O}{\dot{D}_E} \quad (1)$$

$$\dot{D}_O = \frac{\sum_{i=1}^n d_i}{n} \quad (2)$$

$$\dot{D}_E = \frac{0.5}{\sqrt{n/A}} \quad (3)$$

其中， d_i 是点 i 距其最近点的直线距离。 A 的取值为坐标集合构成的不规则区域的外接矩形面积。 $\dot{D}_O$ 是每个坐标点与其最近邻坐标点之间的平均观测距离， $\dot{D}_E$ 是所有坐标点处于均匀分布模式下的平均期望距离ANN值的意义如图6所示，当ANN值大于1时，表示坐标均匀地分布在地理空间内，当ANN值小于1时，则表示坐标更偏向于局部聚集。



(a) 均匀分布, ANN>1 (b) 局部聚集分布, ANN<1

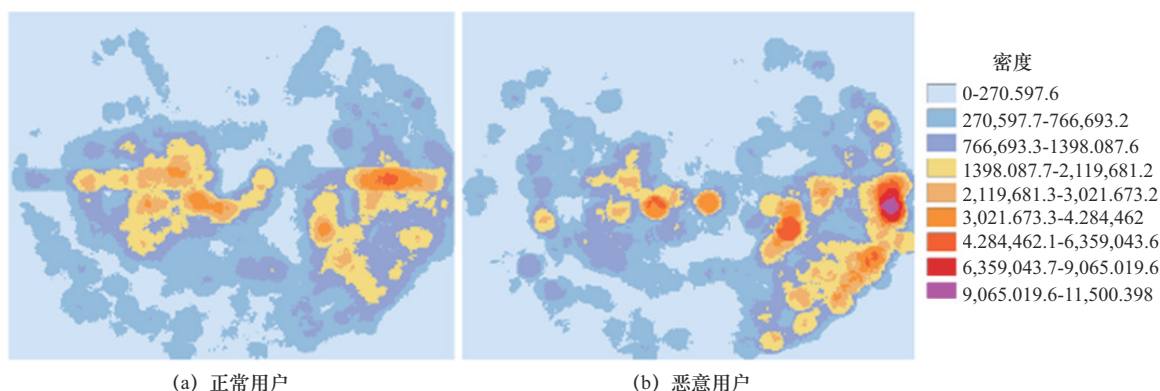
图6 不同空间分布状态对应的ANN值

Figure 6 ANN values corresponding to different spatial distribution states

对比正常用户与恶意用户的历史位置坐标形成的ANN值同样需要平衡双方的用户数量与坐标数量。首先，根据坐标数量对恶意用户群体进行分组，并计算各组内恶意用户的ANN值；接着，对于每一组，根据该组内的恶意用户数量及该组对应的坐标数量区间，在正常用户群体中随机抽取10次相同数量的正常用户，分别计算ANN并求均值。如图7所示，恶意用户在不同坐标数量区间内的ANN值明显小于正常用户。这表明，恶意用户的坐标更偏向于局部聚集（如图6(b)）。即使正常用户的坐标也可能集中在家、公司、常去兴趣点附近，但恶意用户往往是自动化脚本，总是将设备位置模拟到指定坐标，即在一段时间内该用户的坐标总是处于设定值，致使多个坐标重叠于一点，导致ANN值上升。综上所述，本文将通过描述用户历史位置坐标的聚集程度来刻画用户的空间行为特征。

4.3 性别趋向

在社交应用中，最为熟知的一类恶意用户是具有诱惑性内容的女性用户。本文发现在恶意用户集合中，即使该用户没有发布任何内容，其依然更倾向于将性别设置为女性。



(a) 正常用户

(b) 恶意用户

图5 正常用户与恶意用户坐标点密度图

Figure 5 Coordinate point density map of normal and malicious users

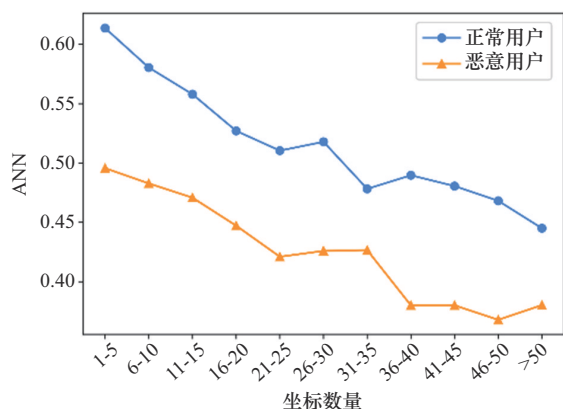


图7 正常用户与恶意用户在不同坐标区间的ANN值

Figure 7 ANN values of normal and malicious users in different coordinate intervals

如图8所示,在正常用户群体与总用户群体中,男性用户占比较女性用户更高,而在恶意用户群体中,女性用户占比高达75.9%。且由于在数据集 $D_{ZhengZhou}$ 中,恶意用户占比仅10.33%,而正常用户占比为89.67%,正常用户的数量远高于恶意用户。因此,正常用户群体中的男女比例对总用户群体中的男女比例起主导作用,总用户群体中依然是男性用户占比更高,与之相对的是恶意用户群体中女性用户占比更高的现象。

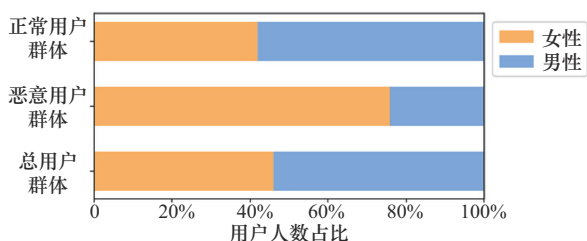


图8 男性与女性用户人数占比

Figure 8 The proportion of male and female users

恶意用户更倾向于将性别设置为女性的这种现象在陌生人交友平台中并不罕见,恶意用户利用男性用户对女性的兴趣和信任,将性别设置为女性有助于建立起表面上的信任关系。随着交流的深入,恶意用户可能会利用男性用户的情感或其他手段进行欺骗,从而达到他们的不法目的。而男性用户常常成为不法活动的受害者。因此,本文将性别作为检测恶意用户的重要特征之一。

5 系统设计

本节基于上文对恶意用户时空行为的分析,

针对性地提出了用户时空行为特征构建方法。图9给出方法的原理框架,其通过自动提取用户时空行为特征和性别属性来判断被测用户是否是恶意用户。

该框架主要由4部分组成,分别是数据搜集层、输入层、时空行为特征提取层和预测层。在数据搜集层中利用3个以上的探针持续搜集目标用户信息;在输入层中将搜集到的用户信息作为输入,包括探针距目标用户的距离、用户在线时间以及用户性别;在特征提取层中,包括空间行为特征提取与时间行为特征提取两个模块,分别对用户历史位置分布的聚集程度与在线时间偏好进行特征提取;最后,聚合用户时空行为特征与性别属性,并将其输入预测层得到用户的状态标签 $\hat{y}_i$ (0表示正常用户,1表示恶意用户)。接下来,本文将详细描述用户时空行为特征构建方法。

5.1 时间行为特征提取

基于上文得出的恶意用户具有差异性的在线时间偏好,本节通过描述用户在不同时间段的在线概率来描述用户细粒度的在线时间偏好。

首先,通过用户的历史在线时间计算总在线时间天数 O ;然后,提取在线时间中与日期无关的部分得到格式化的在线时间集合 $\varepsilon = \{\tilde{l}_1, \tilde{l}_2, \dots, \tilde{l}_n\}$;接着,以 τ (min)为单位将24 h分割为 κ 个时间间隔,使用一个长度为 κ 的数组表示 $L = \{l_0, l_1, \dots, l_\kappa\}$ 。初始状态下 L 中每个元素都为0,计算过程如图10所示,将集合 ε 中的 $\tilde{l}_i (1 \leq i \leq n)$ 映射到 L 中,使对应元素 $l_i (0 \leq i \leq \kappa)$ 加1;最后将 L 中每个元素除以 O 使 l_i 表示为某时间段的在线概率。

在数据搜集过程中,假定探针在 t 时刻请求陌陌附近页面,此时只能证明列表中的用户在 t 时刻在线,在搜集该列表中用户信息的过程中无法感知用户是否在线。设一轮的搜集时间为 s ,在理想情况下, u_i 总是在线,探针每隔时间 s 搜集到一次 u_i 的在线记录。然而,由于LBSD服务随机推荐部分附近用户,存在某次LBSD服务请求不能发现 u_i ,此时每间隔时间 $s + a$ (a 为随机时间长度)搜集到一次 u_i 的在线记录。若 $\tau \leq s + a$ ($a > s$),则在 L 中必定存在某时间间隔区间判定为 u_i 不在线,故 τ 的取值应大于 s 。

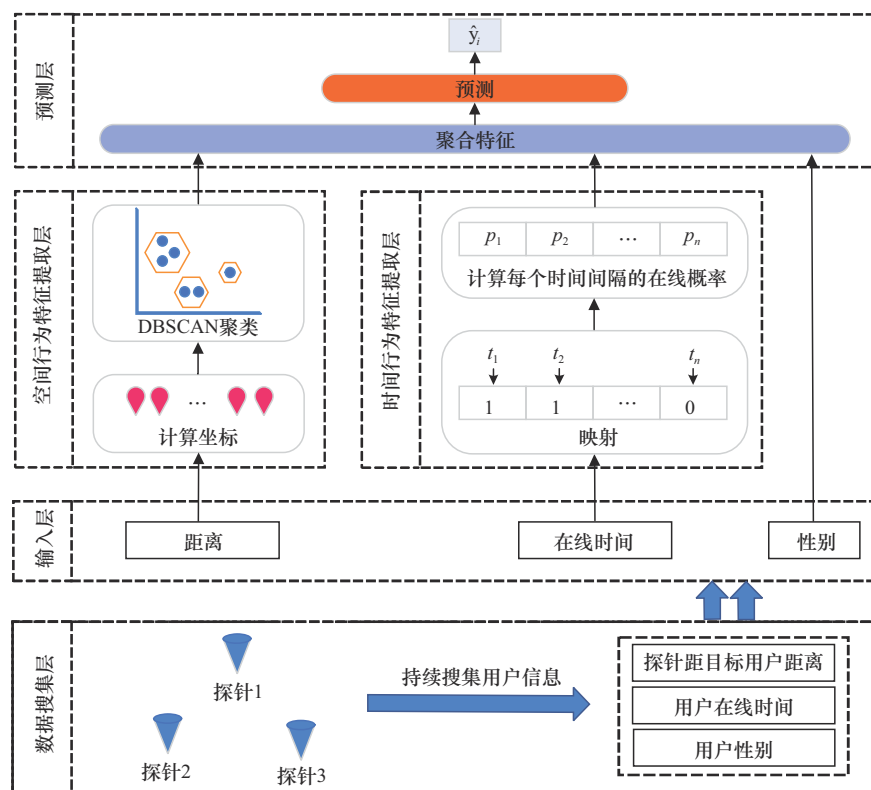


图9 方法的原理框架

Figure 9 The principle framework of the proposed method

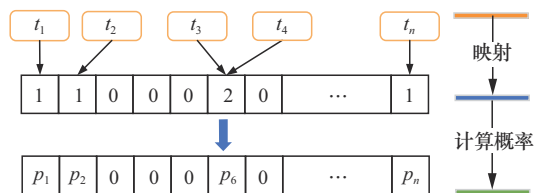


图10 将在线时间映射为不同时间段的在线概率

Figure 10 Map online time to online probabilities for different time periods

由于在数据搜集过程中有6个探针交叉搜集,通过统计数据集得出平均每8.2 min扫描一次陌陌附近页面用户,即 $s = 8.2$ 。本文依据经验依次使用 $\tau = 10$ 、 $\tau = 20$ 、 $\tau = 30$ 、 $\tau = 40$ 对24 h进行分隔,基于数据集 $D_{\text{ZhengZhou}}$ 计算不同类型用户群体总的在线概率均值。如图11(a)、图11(b)、图11(d)中不同类型用户的在线概率曲线表现出了更加分离的趋势;其中图11(a)的 τ 取值过小,导致两种类型用户的概率曲线出现过多重叠,从而影响区分度;而图11(d)中间隔过大,导致过多细节被掩盖;如图11(b)所示,20 min为最佳间隔时间,故 τ 的取值确定为20 min。

5.2 空间行为特征提取

基于上文对用户历史位置坐标分布状态的分析,本节利用DBSCAN^[30]算法对用户的历史位置坐标进行聚类,计算聚集程度。在得到用户的坐标序列后,首先将经纬度坐标系投影至笛卡尔坐标系以便利用聚类算法进行计算。算法的邻域半径参数设置为60 m,这是因为计算陌陌用户坐标的误差在60 m以内,该参数能够确保将用户在同一地点的坐标分类为同一个簇,算法输出的簇个数记为 η 。

图12是对某用户的历史位置坐标集合进行聚类后的结果,图中用六边形标注出了不同的簇,可以看出该用户的坐标集合共形成6个簇。为了统一评估簇内坐标的聚集程度,本文使用Prim^[31]算法计算簇内坐标间最短距离。具体可以表述为簇内坐标点间相互连接,形成一个全连通图,图中每条边的权值为坐标点间的距离。利用Prim算法求得该坐标集合构成的最小生成树的距离 γ ,最后求得所有簇内坐标间距离后取均值得平均簇内坐标间距离 $\bar{\gamma}$ 。

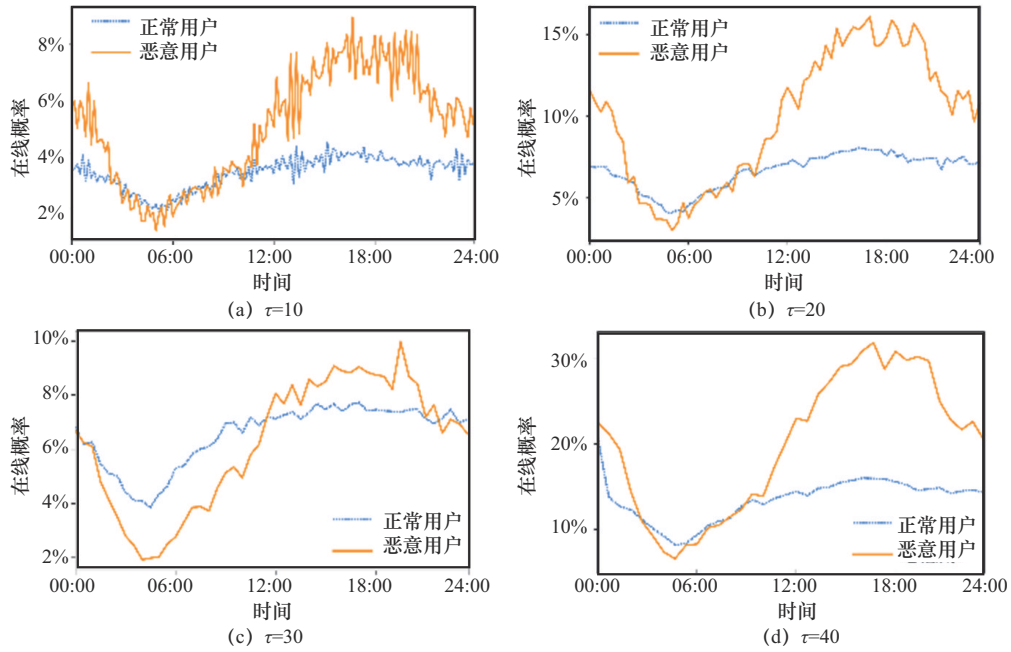
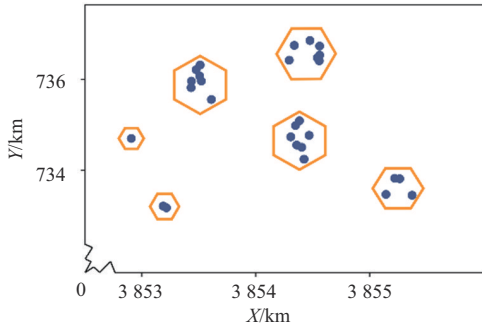
图 11 正常用户群体与恶意用户群体在 τ 不同取值下的平均在线概率Figure 11 The average online probability of normal and malicious user groups at different τ values

图 12 利用 DBSCAN 算法聚类坐标点

Figure 12 Using DBSCAN algorithm to cluster coordinate points

$$\hat{\gamma} = \frac{\sum_{i=1}^{\eta} \gamma_i}{\eta} \quad (4)$$

图 13(a)表示的是数据集 $D_{\text{ZhengZhou}}$ 中所有正常用户与恶意用户的簇数量核密度图,可以看出恶意用户偏向于更少的簇数量,与正常用户的核密度曲线具有一定分离度。其中,当用户 u_i 的记录太少,不足以计算其坐标时,该用户的簇数量为 0。图 13(b)表示的是正常用户与恶意用户的平均簇内坐标间距离核密度图,可以看出恶意用户群体的当平均簇内坐标间距离在小于 20m 时,核密度高于正常用户;当平均簇内坐标间距离大于 20m 时,恶意用户与正常用户的核密度曲线呈现分离状态。在图 13(b)中,坐标间距离为 0 的用户

有两种情况:①用户 u_i 记录太少不足以计算坐标;②用户 u_i 使用模拟位置工具将坐标设置到某位置,此时簇内坐标重合为一点。

本节基于聚类算法描述用户历史位置的聚集程度,以构建用户的空间行为特征。具体为计算用户的历史位置坐标分布形成的簇个数 η 及平均簇内坐标间距离 $\hat{\gamma}$ 。分析结果表明,所提特征能够在一定程度上将恶意用户与正常用户分离。

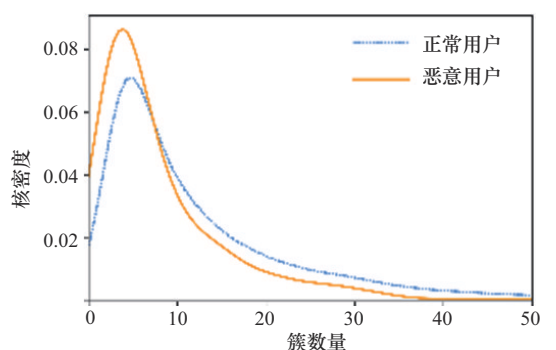
6 实验

6.1 特征描述

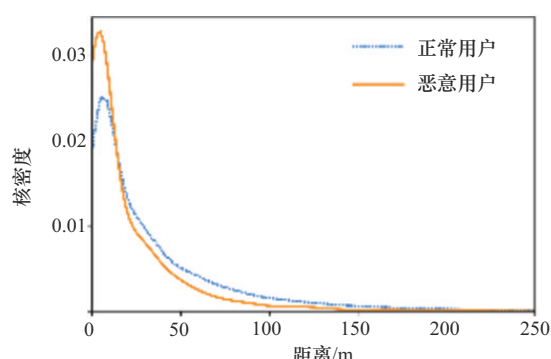
本文通过提取用户在不同时间段的在线概率作为时间特征,并使用聚类算法计算用户历史位置分布的聚集程度作为空间特征。此外,性别也是一个重要特征。综上,所提方法使用到的用户特征如表 4 所示。

6.2 评估指标

本文使用 4 个广泛使用的指标来评估所提方法的性能,分别为受试者工作特征曲线下的面积 (AUC-ROC, area under curve - receiver operating characteristic) [32]、精确率-召回率曲线下的面积 (AUC-PR, area under curve-precision-recall [33])。其中, AUC-ROC 是二分类问题中广泛使用的指标,其值大于 0.5 时说明分类器的分



(a) 不同类型用户的坐标形成的簇数量核密度图



(b) 不同类型用户的平均簇内坐标间距离核密度图

图13 不同类型用户的簇数量与平均簇内坐标间距离核密度图

Figure 13 Kernel density map of the number of clusters for different types of users and the distance between average intra cluster coordinates

表4 所提方法使用到的特征

Table 4 The features used in the proposed method

特征	描述
时间特征	用户在不同时间段的在线概率
空间特征	用户历史位置坐标形成的簇数量与平均簇内坐标间距离
性别	用户的性别

类性能优于随机猜测，值越大代表分类器性能越好。考虑到本文数据集中恶意用户占比仅约11%，恶意用户与正常用户的比例极不平衡，单使用AUC-ROC不足以准确评估不同方法面对这种不平衡数据集的真实性能差距，因此，本文选择AUC-PR作为补充指标。AUC-PR更加关注正样本的预测准确性，取值范围为0~1，值越大代表分类器性能越好。与此同时，本文还使用recall和precision对不同方法进行定量评估，这两个指标的定义分别如式(5)和式(6)所示。

$$\text{recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (6)$$

其中，TP (true positive) 表示正确分类恶意用户的个数；FP (false positive) 表示错误分类恶意用户的个数；FN (false negative) 表示错误分类正常用户的个数。

6.3 验证方法

本文使用十折交叉验证对所提方法的性能进行评估。首先，将整个数据集中的数据进行随机重排序；然后，把数据集平分为10份，将其中的9份数据作为训练集，剩余的1份数据作为测试集；最后，使用评估指标验证被测方法的分类性能。上述过程重复10次，以确保每份数据都有机会被作为测试集，最终取10次验证结果的均值作为最终结果。

6.4 数据集描述

实验共使用了3个数据集，分别是 $D_{\text{ZhengZhou}}$ 、 D_{ShangHai} 和 D_{ChengDu} ，数据集描述如表5所示。其中，发帖数量由用户公开的帖子数量统计得出，而不是用户主页显示的发帖数量。这是因为陌陌应用允许用户对帖子进行隐藏，在这种情况下，无法获取该用户的全部帖子。同时，用户的历史位置坐标由上文提出的用户坐标计算方法得出。

表5 数据集描述

Table 5 Dataset description

数据集属性	$D_{\text{ZhengZhou}}$	D_{ShangHai}	D_{ChengDu}
用户数量/个	39 614	14 582	13 084
恶意用户占比	10.33%	11.92%	10.79%
坐标数量/个	657 378	214 128	169 148
在线时间记录数量/条	1 199 825	393 850	301 242
发帖数量/条	287 844	99 169	87 056
评论数量/条	256 822	89 522	64 892

6.5 实验实施

本节通过具体实验确定分类算法，然后选取现有代表性方法作为对比方法，基于本文数据集对不同方法的性能进行评估。

6.5.1 分类器选择

Han等^[34]对主流的异常检测算法进行了性能评估，其中的在有监督学习算法，CatBoost^[35]、XGBoost^[36]与LightGBM^[37]的性能更加优异。本

节基于数据集 $D_{\text{ZhengZhou}}$ ，并使用以上 3 种算法作为所提方法的分类器进行恶意用户检测，对比性能差异。如表 6 所示，不同算法的性能没有显著差异，这表明所提方法对分类器的选择不敏感。同时，XGBoost 与 LightGBM 的性能略高于 CatBoost，并且 LightGBM 的 AUC-ROC 仅比 XGBoost 低 0.35%，而 AUC-PR 比 XGBoost 高 1.16%。在恶意用户检测工作中，当不同分类器的 AUC-ROC 无显著差异时，AUC-PR 是一个更重要的指标。因此，在本节以下的实验中，所提方法均使用 LightGBM 作为分类器。

表 6 不同分类算法的性能

Table 6 Performance of different classification algorithms

分类算法	AUC-ROC	AUC-PR
CatBoost	0.850	0.599
XGBoost	0.858	0.606
LightGBM	0.855	0.613

6.5.2 与现有方法对比分析

本节选择两个代表性方法进行对比分析实验，分别是基于社交关系图的方法 SybilSCAR^[12]和基于用户发布内容的方法 DatingSec^[17]，本文遵循它们的默认设置。其中，DatingSec 对非活跃用户的定义与先前针对传统社交网络的恶意用户检测方法^[38]相似，其将有用信息（发帖、评论等）较少的用户定义为非活跃用户，并且不考虑这些“非活跃用户”，理由是这些“非活跃用户”可用信息不足，对这部分用户进行检测存在困难。DatingSec 使用这种策略剔除了其数据集中发帖数量小于等于 5 的“非活跃用户”，需要注意的是这部分用户在其数据集中占比约 65%，应当考虑对这部分用户进行检测。

图 14 展示了本文数据集中恶意用户的发帖数量统计，在所有数据集中，超过 70% 的恶意用户发帖数量在 10 条以内，并且恶意用户发帖数量的中位数不高于 6。因此，发帖数量较少的恶意用户在总恶意用户集合中属于多数群体，忽视发帖数量较少的用户可能会造成大量恶意用户逃脱检测机制。因此，本文重点关注这些发帖较少的恶意用户群体，并将所提方法与现有代表性方

法进行对比分析。

对于数据集 $D_{\text{ZhengZhou}}$ ，表 7 展示了仅使用该数据集中发帖数量小于等于 5 的恶意用户时不同方法的 recall 和 precision 值，并且不同方法判定为恶意用户的概率阈值均设定为 0.5。可以看到基于社交关系图的方法 SybilSCAR 性能较低，这是因为陌陌应用以陌生人交友为主要功能，用户间联系松散。同时，相比 DatingSec，所提方法在损失少量 precision 的情况下，recall 提升 50.82%，即所提方法发现恶意用户的能力显著高于 SybilSCAR 和 DatingSec。

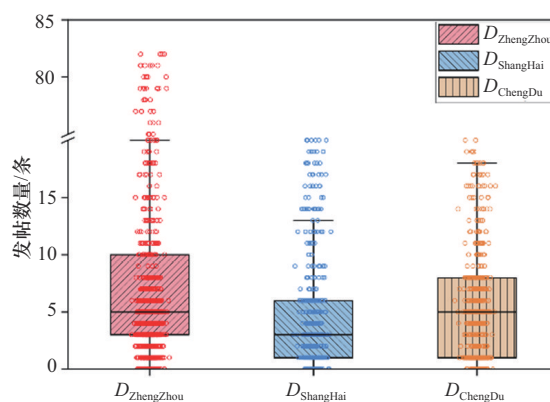


图 14 不同数据集中恶意用户发帖数量统计

Figure 14 Statistics on the number of malicious user posts in different datasets

表 7 仅使用发帖数量小于等于 5 的恶意用户时不同方法的检测结果

Table 7 Detection results of different methods when using only malicious users with a number of posts less than or equal to 5

方法	recall	precision
所提方法	0.825	0.504
DatingSec ^[23]	0.547	0.560
SybilSCAR ^[18]	0.336	0.298

图 15 与图 16 展示了将不同方法应用在数据集 $D_{\text{ZhengZhou}}$ 上得出的 AUC-ROC 与 AUC-PR 值。可以看到与所提方法相比，随着恶意用户发帖数量的减少，DatingSec 和 SybilSCAR 的 AUC-ROC 与 AUC-PR 均大幅度降低，这说明用户可用信息过少可能会导致现有方法检测性能不稳定。而本文所提方法不受上述情况影响，当只使用发帖数量小于等于 5 的恶意用户时，所提方法与 Dating-

Sec 和 SybilSCAR 相比, AUC-ROC 分别提升 4.63% 和 40.03%, AUC-PR 分别提升 19.88% 和 198.48%。更重要的是, 在恶意用户检测工作中, 普遍关注发现恶意用户的能力, 为了避免遗漏, 将少量的正常用户误判为恶意用户不可避免。所提方法的 AUC-PR 值更高表明, 所提方法能够更好地平衡发现恶意用户的能力与误判率, 即与 DatingSec 和 SybilSCAR 相比, 所提方法在发现更多恶意用户的同时能够减少对正常用户的误判。

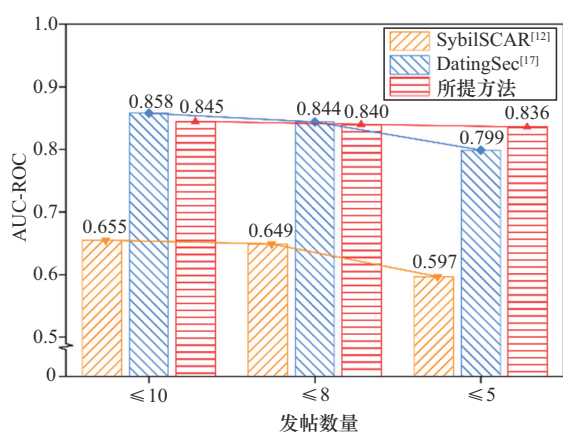


图15 根据恶意用户发帖数量进行分组后不同方法的 AUC-ROC 值

Figure 15 AUC-ROC values for different methods after grouping based on the number of malicious user posts

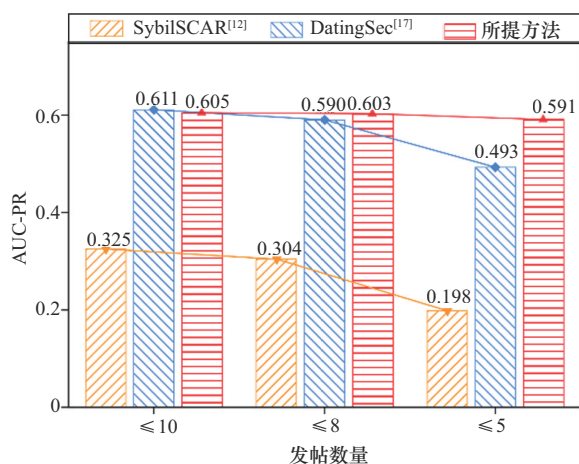


图16 根据恶意用户发帖数量进行分组后不同方法的 AUC-PR 值

Figure 16 AUC-PR values for different methods after grouping based on the number of malicious user posts

为了验证所提方法的泛化能力, 下述实验将 $D_{ZhengZhou}$ 作为训练集, 在该集合中, 用户是否恶

意用户的标签已知; 将数据集 $D_{ShangHai}$ 与 $D_{ChengDu}$ 作为验证集, 该集合中, 用户是否恶意用户的标签未知。并且仅使用发帖数量小于等于 5 的恶意用户, 以验证不同方法在用户可用信息有限情况下的检测性能。由于训练集与验证集被明确划分, 以下实验不再使用十折交叉验证。

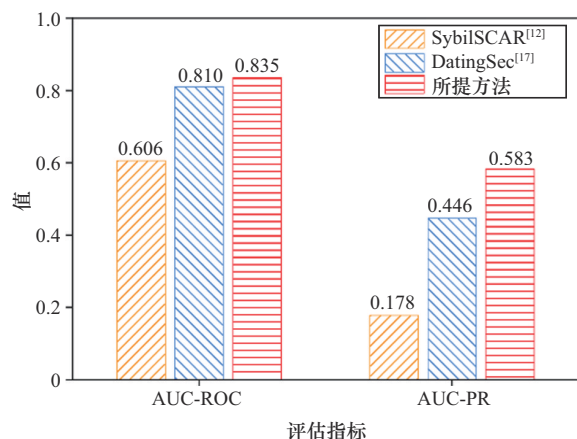
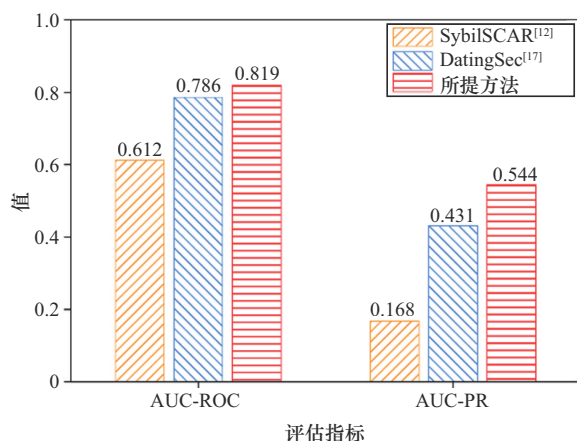
表 8 展示了不同方法在数据集 $D_{ShangHai}$ 与 $D_{ChengDu}$ 的检测结果。其中, 不同方法判定为恶意用户的概率阈值均设定为 0.5。可以看到所提方法以损失少量 precision 为代价获得了最高的 recall 值, 在两个数据集上的 recall 值显著高于 DatingSec 和 SybilSCAR。固定阈值下更高的召回率说明所提方法对恶意用户的检测覆盖率更高, 但在实际应用中, 应当根据需求制定合理的阈值。因此, 以下实验将使用 AUC-ROC 与 AUC-PR 衡量不同方法在各个阈值下的综合性能。

表8 不同方法在数据集 $D_{ShangHai}$ 与 $D_{ChengDu}$ 上的检测结果

Table 8 Detection results of different methods on datasets $D_{ShangHai}$ and $D_{ChengDu}$

方法	$D_{ShangHai}$		$D_{ChengDu}$	
	recall	precision	recall	precision
所提方法	0.811	0.473	0.805	0.498
DatingSec ^[23]	0.544	0.490	0.602	0.535
SybilSCAR ^[18]	0.266	0.282	0.353	0.330

如图 17 与图 18 所示, 所提方法在两个数据集上均表现出了最好的性能。将数据集 $D_{ShangHai}$ 作为验证集时, 所提方法的 AUC-ROC 与 AUC-PR 分别高于对比方法 3.09%、30.72% 以上; 将数据集 $D_{ChengDu}$ 作为验证集时, 所提方法的 AUC-ROC 与 AUC-PR 分别高于对比方法 4.2% 和 26.22% 以上。这表明所提方法在面对未知数据集时依然能够有效检测出潜在的恶意用户, 具有一定的泛化能力。另外, 从图 17 和图 18 可以看出, 所提方法在数据集 $D_{ChengDu}$ 上的性能略低于数据集 $D_{ShangHai}$ 。这可能是数据集 $D_{ChengDu}$ 中用户的人均在线时间记录数量和人均坐标数量分别低于数据集 $D_{ShangHai}$ 14.76% 和 11.96%, 导致性能轻微下降。

图 17 $D_{ShangHai}$ 作为验证集的检测结果Figure 17 Detection results using $D_{ShangHai}$ as the validation set图 18 $D_{ChengDu}$ 作为验证集的检测结果Figure 18 Detection results using $D_{ChengDu}$ as the validation set

与现有方法不同, 所提方法不关注用户的发布内容或与其他用户的交互行为, 仅利用 LBSD 服务提供的少量用户信息即可有效检测出恶意用户。这是因为恶意用户将利用 LBSD 服务进行犯罪活动视为一种工作, 而不是休息之余的消遣。因此, 他们的在线活动规律与正常用户具有一定差异。同时, 恶意用户想要跟踪某个目标区域内的用户, 必须将自己的位置模拟到该区域附近, 而这种模拟的位置很难与正常用户的位置分布特征相似。最后, 本文将正常用户和恶意用户的时空行为与性别属性进行了对比分析, 以确保所提特征均能够对检测结果起到积极影响。

6.5.3 消融实验

本节基于数据集 $D_{ZhengZhou}$ 分析所提方法中的不同特征对检测性能的影响。表 9 展示了使用不同特征时的检测结果, 可以看到空间特征在两项

评估指标上略低于时间特征, 这是由于空间特征虽然能够在一定程度上将正常用户群体与恶意用户群体分离, 但区分度不够充分, 如恶意用户的历史位置坐标形成的簇数量与平均簇内坐标间距离这两种空间行为属性值普遍偏低, 这也是识别恶意用户的一个重要决策点, 但正常用户中也同样存在空间行为属性值较低的用户, 这部分空间行为属性值较低的正常用户与恶意用户相互重叠, 对预测结果造成了干扰。而时间特征表现出了较空间特征更好的性能, 这是由于恶意用户与正常用户的在线时间偏好具有更好的分离度, 仅使用时间特征也能够较有效地检测出潜在的恶意用户。特别地, 当把时间特征与空间特征融合为时空特征时, 两项评估指标较单独使用时间或空间特征均有显著提升, 这说明时间特征与空间特征对预测结果存在一定的互补性, 相互增强, 当目标用户的空间行为属性值较低时, 可增大其是恶意用户的概率, 然后结合时间特征辅助决策, 能够实现更高的检测准确率。

表 9 不同特征对检测性能的影响

Table 9 The effect of different features to detection performance

特征	AUC-ROC	AUC-PR
时间特征	0.729	0.354
空间特征	0.680	0.299
时空特征	0.775	0.472
性别	0.747	0.256
聚合所有特征	0.855	0.613

值得注意的是, 性别属性对应的 AUC-ROC 值相对较高, 仅低于时空特征 3.61%, 但 AUC-PR 值却较时空特征低 43.86%。这是由于单使用性别特征时, 分类器将过多的正常用户预测为了恶意用户。正如数据集 $D_{ZhengZhou}$ 中恶意用户 10.33% 的占比情况, 假如将 10% 的正常用户误判为了恶意用户, 此时对 AUC-ROC 的假正率影响较小, 但对 AUC-PR 中的精确率影响较大, 这会导致较低的 AUC-PR 值。因此, 为了增强恶意用户检测模型的分类能力, 性别可以作为一个辅助特征, 与其他特征结合使用以增强模型的分类能力。当聚合所有特征时, 判断一个用户是否恶

意用户就需综合考虑多个因素, 此时 AUC-ROC 和 AUC-PR 的值均有显著提升。

本节通过消融实验对本文所提特征的重要性进行了分析。结果表明, 所提特征均对检测结果起到了积极影响, 相较单个特征, 将多个特征融合能够显著提升恶意用户检测准确率。

7 结束语

陌陌应用中存在的恶意用户对正常用户构成了威胁。本文检测陌陌应用中的恶意用户, 基于自主搜集的 67 280 个用户的公开资料数据, 发现了恶意用户的时空行为差异化特征, 并设计了基于时空行为与静态属性分析的恶意用户检测方法。实验结果表明, 所提方法能够有效发现陌陌应用中的恶意用户, 尤其在用户历史发布内容较少的情况下, 所提方法对恶意用户的检测性能明显优于现有的具有代表性的方法。在未来的工作中, 将研究如何融合用户间社交关系、时空行为和发布内容等特征, 结合深度学习方法进一步提升恶意用户检测的准确率。

参考文献:

- [1] COBB C, KOHNO T. How public is my private life privacy in online dating[C]//Proceedings of the 26th International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017: 1231-140.
- [2] CHENG Z Y, CAVERLEE J, LEE K, et al. Exploring millions of footprints in location sharing services[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2021, 5(1): 81-88.
- [3] LI J, YAN H Y, LIU Z L, et al. Location-sharing systems with enhanced privacy in mobile online social networks[J]. IEEE Systems Journal, 2017, 11(2): 439-448.
- [4] CHEN T, KAAAFAR M, BORELI R. The where and when of finding new friends: analysis of a location-based social discovery network [J]. Proceedings of the International AAAI Conference on Web and Social Media, 2021, 7(1): 61-70.
- [5] LI H X, ZHU H J, DU S G, et al. Privacy leakage of location sharing in mobile social networks: attacks and defense[J]. IEEE Transactions on Dependable and Secure Computing, 2018, 15(4): 646-660.
- [6] THILAKARATHNA K, SENEVIRATNE S, GUPTA K, et al. A deep dive into location-based communities in social discovery networks [J]. Computer Communications, 2017, 100: 78-90.
- [7] HONG G, YANG Z M, YANG S, et al. Analyzing ground-truth data of mobile gambling scams[C]//Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2022: 2176-2193.
- [8] DING Y, PEDDINTI S T, ROSS K W. Stalking Beijing from Timbaktu: a generic measurement approach for exploiting location-based social discovery[C]//Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices. New York: ACM, 2014: 75-80.
- [9] WANG J L, CHENG H N, XUE M H, et al. Revisiting localization attacks in mobile app people-nearby services[C]//International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage. Cham: Springer, 2017: 17-30.
- [10] CAO Q, SIRIVIANOS M, YANG X W, et al. Aiding the detection of fake accounts in large scale social online services[C]//Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation. New York: ACM, 2012: 15.
- [11] GONG N Z, FRANK M, MITTAL P. SybilBelief: a semi-supervised learning approach for structure-based sybil detection[J]. IEEE Transactions on Information Forensics and Security, 2014, 9(6): 976-987.
- [12] WANG B H, ZHANG L, GONG N Z. SybilSCAR: Sybil detection in online social networks via local rule based propagation[C]//Proceedings of the IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.
- [13] GAO P, WANG B H, GONG N Z, et al. SYBILFUSE: combining local attributes with global structure to perform robust sybil detection[C]//Proceedings of the 2018 IEEE Conference on Communications and Network Security (CNS). Piscataway: IEEE Press, 2018: 1-9.
- [14] LI S Y, YANG J, LIANG G, et al. SybilFlyover: Heterogeneous graph-based fake account detection model on social networks[J]. Knowledge-Based Systems, 2022, 258: 110038.
- [15] WANG B H, JIA J Y, GONG N Z. Graph-based security and privacy analytics via collective classification with joint weight learning and propagation[C]//Proceedings 2019 Network and Distributed System Security Symposium. Reston, VA: Internet Society, 2019: 24-27.
- [16] AL-QURISHI M, HOSSAIN M S, ALRUBAIAN M, et al. Leveraging analysis of user behavior to identify malicious activities in large-scale social networks[J]. IEEE Transactions on Industrial Informatics, 2018, 14(2): 799-813.
- [17] HE X L, GONG Q Y, CHEN Y, et al. DatingSec: detecting malicious accounts in dating Apps using a content-based attention network[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(5): 2193-2208.
- [18] SHEN X F, LV W T, QIU J H, et al. Trust-aware detection of malicious users in dating social networks[J]. IEEE Transactions on Computational Social Systems, 2023, 10(5): 2587-2598.
- [19] BREUER A, KHOSRAVANI N, TINGLEY M, et al. Preemptive detection of fake accounts on social networks via multi-class preferential attachment classifiers[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2023: 105-116.
- [20] XU T, GOOSSEN G, CEVAHIR H K, et al. Deep entity classification: abusive account detection for online social networks[C]//USENIX Security Symposium. [S.l.:s.n.], 2021: 4097-4114.
- [21] SAEED M H, ALI S, BLACKBURN J, et al. TrollMagnifier: de-

- protecting state-sponsored troll accounts on reddit[C]//Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2022: 2161-2175.
- [22] ZHU H Y, WANG C. Enabling fraud prediction on preliminary data through information density booster[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 5706-5720.
- [23] LI M Y, ZHU H J, GAO Z Y, et al. All your location are belong to us: breaking mobile social networks for automated user location tracking[C]//Proceedings of the 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM, 2014: 43-52.
- [24] CHENG H N, MAO S L, XUE M H, et al. On the impact of location errors on localization attacks in location-based social network services[C]//WANG G, RAY I, ALCARAZ CALERO J, et al. International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage. Cham: Springer, 2016: 343-357.
- [25] SHI W Q, LUO X Y, ZHAO F, et al. Geolocating a WeChat user based on the relation between reported and actual distance[J]. International Journal of Distributed Sensor Networks, 2018, 14(4): 1550147718774462.
- [26] PENG J W, MENG Y, XUE M H, et al. Attacks and defenses in location-based social networks: a heuristic number theory approach [C]//Proceedings of the 2015 International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec). Piscataway: IEEE Press, 2015: 64-71.
- [27] POLAKIS I, ARGYROS G, PETSIOS T, et al. Where's wally? precise user discovery attacks in location proximity services[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2015: 817-828.
- [28] 时文旗, 罗向阳, 郭家山. 基于加权最小二乘的社交网络用户定位方法[J]. 网络与信息安全学报, 2022, 8(3): 41-52.
- SHI W Q, LUO X Y, GUO J S. Social network user geolocating method based on weighted least squares[J]. Chinese Journal of Network and Information Security, 2022, 8(3): 41-52.
- [29] ANDY M, LAUREN S G. The ESRI guide to GIS analysis, volume 2: spatial measurements and statistics[M]. Redlands: ESRI press, 2005: 109-111.
- [30] ESTER M, KRIEDEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. New York: ACM, 1996: 226-231.
- [31] PRIM R C. Shortest connection networks and some generalizations [J]. Bell System Technical Journal, 1957, 36(6): 1389-1401.
- [32] FAWCETT T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.
- [33] DAVIS J, GOADRIC M. The relationship between Precision-Recall and ROC curves[C]//Proceedings of the 23rd International Conference on Machine learning - ICML '06. New York: ACM, 2006: 233-240.
- [34] HAN S Q, HU X Y, HUANG H L, et al. ADBench: anomaly detection benchmark[EB]. arXiv Preprint, 2022, arXiv: 2206.09426.
- [35] PROKHORENKOVA L, GUSEV G, VOROBIEV A, et al. CatBoost: unbiased boosting with categorical features[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 6639-6649.
- [36] 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, 2016: 785 - 794.
- [37] KE G L, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 3149-3157.
- [38] YANG Z, WILSON C, WANG X, et al. Uncovering social network Sybils in the wild[J]. ACM Transactions on Knowledge Discovery from Data, 2014, 8(1): 2.

[作者简介]



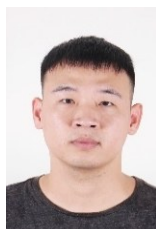
杨海峰 (1997-), 男, 河南安阳人, 郑州大学硕士生, 主要研究方向为移动社交应用隐私安全、网络空间资源测绘。



杜少勇 (1989-), 男, 河北任县人, 博士, 河南省网络空间态势感知重点实验室讲师, 主要研究方向为移动安全。



魏果 (1997-), 女, 河南信阳人, 河南省网络空间态势感知重点实验室博士生, 主要研究方向为网络空间资源测绘。



时文旗 (1993-), 男, 河南周口人, 博士, 河南省网络空间态势感知重点实验室讲师, 主要研究方向为网络目标定位、网络空间资源测绘。



罗向阳 (1978-), 男, 湖北荆门人, 河南省网络空间态势感知重点实验室教授、博士生导师, 主要研究方向为网络空间资源测绘、多媒体安全。