

新的基于鼠标行为的持续身份认证方法

易聪^{1,2}, 胡军^{1,2}

(1. 计算智能重庆市重点实验室(重庆邮电大学), 重庆 400065; 2. 重庆邮电大学计算机科学与技术学院, 重庆 400065)

摘要: 随着互联网技术的快速发展, 安全问题一直是人们关注的焦点。基于鼠标行为的持续身份认证对于保护计算机系统的安全起着至关重要的作用。针对鼠标行为认证方法存在的认证准确率较低和认证时间较长的问题, 提出了一种新的基于鼠标行为的持续身份认证方法。该方法将用户的鼠标事件序列按不同的类型划分为相应的鼠标行为, 并基于鼠标行为从多方面挖掘鼠标行为特征, 对于时间和空间的特征值需要计算其统计值来唯一表示, 从而更好地表示不同用户的鼠标行为差异, 提高认证准确率。通过 ReliefF 算法得到鼠标行为特征的重要度, 并在此基础上结合邻域粗糙集去除鼠标行为的无关或冗余特征, 以达到降低模型复杂度和建模时间的目的, 并采用二分类算法进行认证模型的训练。在身份认证时会根据每次收集的鼠标行为使用认证模型得到分类得分, 再结合信任模型更新用户的信任值, 当用户的信任值降低到信任模型阈值以下时, 就会被判断为非法用户。在 Balabit 和 DFL 数据集上对所提方法的身份认证效果进行仿真实验, 结果表明, 该方法相较于其他文献的方法, 不仅可以提高身份认证准确率、降低身份认证时间, 而且对于外部用户的非法入侵具有一定的鲁棒性。

关键词: 身份认证; 鼠标行为; 邻域粗糙集; 特征选择; 信任模型

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2022067

Novel continuous identity authentication method based on mouse behavior

YI Cong^{1,2}, HU Jun^{1,2}

1. Chongqing Key Laboratory of Computational Intelligence(Chongqing University of Posts and Telecommunications), Chongqing 400065, China

2. College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: With the rapid development of Internet technologies, security issues have always been the hot topics.

收稿日期: 2022-01-09; **修回日期:** 2022-06-22

通信作者: 胡军, hujun@cqupt.edu.cn

基金项目: 国家自然科学基金(61936001, 61876201, 61876027); 重庆市教委重点合作项目(HZ2021008); 重庆市自然科学基金(cstc2019jcyj-cxttX0002, cstc2021ycjh-bgzxm0013)

Foundation Items: The National Natural Science Foundation of China (61936001, 61876201, 61876027), The Key Cooperation Project of Chongqing Municipal Education Commission (HZ2021008), The National Natural Science Foundation of Chongqing (cstc2019jcyj-cxttX0002, cstc2021ycjh-bgzxm0013)

引用格式: 易聪, 胡军. 新的基于鼠标行为的持续身份认证方法[J]. 网络与信息安全学报, 2022, 8(5): 179-188.

Citation Format: YI C, HU J. Novel continuous identity authentication method based on mouse behavior[J]. Chinese Journal of Network and Information Security, 2022, 8(5): 179-188.

Continuous identity authentication based on mouse behavior plays a crucial role in protecting computer systems, but there are still some problems to be solved. Aiming at the problems of low authentication accuracy and long authentication latency in mouse behavior authentication method, a new continuous identity authentication method based on mouse behavior was proposed. The method divided the user's mouse event sequence into corresponding mouse behaviors according to different types, and mined mouse behavior characteristics from various aspects based on mouse behaviors. Thereby, the differences in mouse behavior of different users can be better represented, and the authentication accuracy can be improved. Besides, the importance of mouse behavior features was obtained by the ReliefF algorithm, and on this basis, the irrelevant or redundant features of mouse behavior were removed by combining the neighborhood rough set to reduce model complexity and modeling time. Moreover binary classification was adopted. The algorithm performed the training of the authentication model. During identity authentication, the authentication model was used to obtain a classification score based on the mouse behavior collected each time, and then the user's trust value was updated in combination with the trust model. When the user's trust value fell below the threshold of the trust model, it might be judged as illegal user. The authentication effect of the proposed method was simulated on the Balabit and DFL datasets. The results show that, compared with the methods in other literatures, this method not only improves the authentication accuracy and reduces the authentication latency, but also has a certain robustness to the illegal intrusion of external users.

Keywords: identity authentication, mouse behavior, neighborhood rough set, feature selection, trust model

0 引言

在互联网的快速发展下, 计算机技术得到了广泛普及, 越来越多的人需要通过计算机访问各种系统或者应用, 身份认证在其中扮演着“门禁”的角色, 能够判断当前使用计算机系统的用户身份^[1]。现有的身份认证主要有口令认证、持有物认证和生物特征认证。口令认证如账户密码等, 需要人们记忆, 并且容易被盗取、破解; 持有物认证如小区的门禁卡等, 需要人们随身携带, 使用不方便且容易丢失; 基于生物特征的认证方法相较于前两种认证方法, 在安全性和使用性上进行了较大的提升, 成为身份认证的研究热点。

基于生物特征的认证方法主要分为基于生理特征的认证方法和基于行为特征的认证方法。基于生理特征的身份认证方法主要有指纹识别^[2]、人脸识别^[3]、虹膜识别^[4]等。这些认证方法因为需要额外的硬件设备, 所以需要付出较高的成本开销。基于行为特征的认证方法不需要另外的数据收集设备, 仅使用用户与计算机进行交互的行为数据, 成本代价较低, 且这类认证方法可以方便地部署在计算机上, 主要用于计算机系统的持续身份认证, 已经成为身

份认证重要的发展方向^[5-15]。

击键和鼠标动力学是两种主要的基于行为特征的身份认证方法。先前已经有大量研究人员对基于击键行为的身认证方法开展了较为深入的研究^[5-6]。但是击键行为可能包含用户的敏感数据, 如账户、密码等, 而鼠标行为不会泄露用户的隐私, 相对更安全, 因此得到研究人员的广泛关注。

Ahmed 等^[7-8]证明了不同用户的鼠标行为特征存在较大区别, 能够用来识别用户的身份信息。Zheng 等^[9-10]提出了一些基于角度的细粒度特征, 这些新指标相对独特, 独立于计算平台, 并使用 SVM 作为认证模型。徐剑等^[11]采用层次划分的方法对用户鼠标行为进行定义, 并使用随机森林作为分类器以解决数据过拟合和噪声问题。Yildirim 等^[12]能在每个会话中将原始数据划分为鼠标移动, 然后根据会话时间大小对鼠标移动进行分类, 并通过计算长短鼠标移动的独立属性为每个用户创建单独的签名, 最后使用二分类器进行身份认证。田杰等^[13]针对当前基于鼠标行为的身认证方法在自由环境中认证性能低的问题, 提出了一种联合鼠标及窗口行为进行跨域分析的方案。Chong 等^[14]使用卷积神经网络训练用户的鼠标移动轨迹图像, 平均 AUC (ROC 曲线下的面积)

最高能达到 96%。

根据相关经验提取的鼠标行为特征存在一些无关或冗余特征，不仅会增加模型的复杂度，还会增加建模时间。沈超等^[15]提出顺序前向特征选择的方法，需要计算每个特征组合所对应的认证准确率。该方法实现简单但适用性不好，当特征数量增加时，计算时间将大幅度增加。为此，本文结合运行效率高的 ReliefF 算法和邻域粗糙集进行特征选择，在减少特征选择的时间复杂度的同时，能有效去除无关或冗余特征。现有的鼠标行为身份认证通常是通过收集固定数量的鼠标行为数据做出判断，但每位用户都会受到环境的影响，造成鼠标操作行为发生变异，从而无法确定固定收集多少个鼠标行为才能正确判断出该用户的身份，同时如果使用较多数量的鼠标行为判断用户的身份，会造成身份认证时间较长，那么计算机系统的风险将会增大。为此，本文结合信任模型持续地判断用户鼠标行为的合法性，减少误判率，同时由于收集更少数量的鼠标行为就能判断出非法用户，身份认证时间也能降低。

1 基于鼠标行为的身份认证

1.1 鼠标行为划分

鼠标行为是由一系列的鼠标事件组成的。具体地，鼠标行为有以下几种类型。

1) 鼠标移动行为：表示一般的鼠标移动操作，这种类型的操作只需在屏幕的两点之间移动鼠标指针。

2) 鼠标点击行为：表示单击操作，这类类似于鼠标移动行为，但它以鼠标单击事件结束。

3) 鼠标拖拽行为：表示拖拽操作，总是以鼠标左键按下事件开始，然后是一系列鼠标拖拽事件，最后以鼠标左键释放事件结束。

4) 未识别鼠标行为：表示不能划分为以上鼠标行为时为未识别行为。

鼠标行为数量与身份认证准确率的关系如图 1 所示，从图中可以看出，随着身份认证时所使用的鼠标行为数量的上升，认证准确率也上升，但是考虑认证时间与认证准确率的平衡，本文选取行为数量为 10 来做实验。

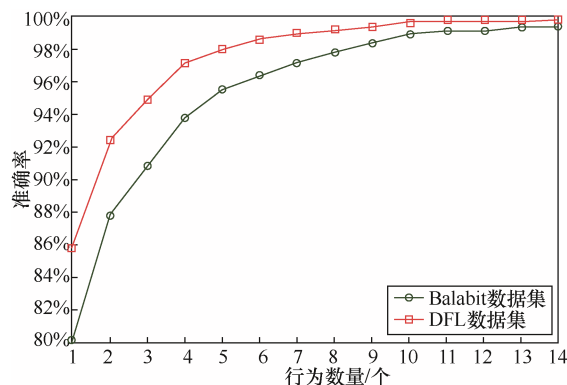


图1 鼠标行为数量与身份认证准确率的关系

Figure 1 The relationship between the number of mouse behaviors and the accuracy of identity authentication

1.2 鼠标行为特征提取

本文将第 i 个时刻采集的鼠标事件表示为 $\{t_i, b_i, st_i, x_i, y_i\}$ ， t_i 为采集时刻， b_i 和 st_i 分别为鼠标按钮状态以及附加信息， x_i 与 y_i 为鼠标指针在屏幕上的坐标。

鼠标事件序列是由原始鼠标事件组成的，通过用户鼠标事件序列可以得到 3 个基础向量。

1) x ：横坐标向量，由鼠标事件 x_i 的值组成，其中 $i=1, \dots, n$ 。

2) y ：纵坐标向量，由鼠标事件 y_i 的值组成，其中 $i=1, \dots, n$ 。

3) t ：时间向量，由鼠标事件 t_i 的值组成，其中 $i=1, \dots, n$ 。

得到以上基础向量后，可通过相关的公式计算出鼠标行为特征值，包括时间上的、空间上的以及其他主要特征值。鼠标事件序列中两相邻鼠标事件 x, y, t 的间隔表示方式分别如下。

$$\delta x_i = x_{i+1} - x_i \quad (1)$$

$$\delta y_i = y_{i+1} - y_i \quad (2)$$

$$\delta t_i = t_{i+1} - t_i \quad (3)$$

(1) 时间上的特征值

$$1) v_x: \text{水平方向的速度}, v_x = \frac{\delta x}{\delta t}$$

$$2) v_y: \text{垂直方向的速度}, v_y = \frac{\delta y}{\delta t}$$

$$3) v: \text{切向方向的速度}, v = \sqrt{v_x^2 + v_y^2}$$

$$4) a: \text{切向方向的加速度}, a = \frac{\delta v}{\delta t}$$

5) j : 切向方向的加加速度, $j = \frac{\delta_a}{\delta_t}$

6) ω : 角速度, $\omega = \frac{\delta_\theta}{\delta_t}$

(2) 空间上的特征值

1) θ : 移动角度, $\theta = \text{atan}2\left(\frac{\delta_y}{\delta_x}\right)$

2) c : 曲率, $c = \frac{\delta_\theta}{\delta_s}$

3) Δc : 曲率变化率, $\Delta c = \frac{\delta_c}{\delta_s}$

(3) 其他主要特征值

1) T : 时间, 表示鼠标行为从起点到终点的时间间隔。

$$T = t_n - t_1$$

2) trajectory length: 轨迹长度, 表示鼠标行为从起点到终点所移动的路径长度。

$$\text{trajectory length} = \sum_{i=1}^{n-1} \sqrt{\delta x_i^2 + \delta y_i^2}$$

3) S : 起点到终点的直线距离。

$$S = \sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2}$$

4) largest_deviation: 最大偏离。

$$\text{largest_deviation} = \max_i \{d(P_i, |P_1 P_n|)\}$$

其中, $i = 1, \dots, n$

5) straightness: 平直度。

$$\text{straightness} = \frac{S}{\text{trajectory length}}$$

6) CP: 关键点数。

$$\text{CP} = \sum_{i=1}^n Z_i, Z_i = \begin{cases} 1, \Delta c_i = 0 \text{ 且 } |c_i| > \alpha \\ 0, \text{ 其他} \end{cases}$$

其中, $\alpha > \frac{\pi}{10} \text{rad/pixel}^2$ 。

根据以上特征计算公式, 最终得到的鼠标行为特征如表1所示。

1.3 鼠标行为特征选择

特征选择是从一组特征中选择对分类结果最有效的特征以降低特征空间维数的过程^[16]。特征选择不仅可以降低模型的复杂度, 还可以减少模型的训练时间。本文提取了大量的鼠标行为特征, 但其中存在一些冗余或无关特征, 需要进行特征选择。考虑到鼠标行为特征都是连续型数据, 而邻域粗糙集能较好地处理连续型数据的特征选择, 可以用来处理鼠标行为特征的选择。

为了能够得到尽可能准确的特征子集, 基于邻域粗糙集的特征选择算法通常需要在每次特征选择时计算所有未选择特征的重要度, 而重要度

表1 鼠标行为特征
Table 1 Mouse behavior features

特征	描述	数量
横坐标	横坐标的最小值、最大值、标准差、平均值、偏度、峰度	6
纵坐标	纵坐标的最小值、最大值、标准差、平均值、偏度、峰度	6
速度	切线, 水平, 竖直方向速度的最小值、最大值、标准差、平均值、偏度、峰度	18
加速度	加速度的最小值、最大值、标准差、平均值、偏度、峰度	6
加加速度	加加速度的最小值、最大值、标准差、平均值、偏度、峰度	6
角度	角度的最小值、最大值、标准差、平均值、偏度、峰度、和	7
角速度	角速度的最小值、最大值、标准差、平均值、偏度、峰度	6
曲率及曲率变化率	曲率及曲率变化率的最小值、最大值、标准差、平均值、偏度、峰度	12
时间间隔, 停顿次数, 停顿时间, 停顿时间比例, 鼠标事件数量, 关键点数, 点击鼠标时间	各1个	7
轨迹长度, 直线距离, 最大偏离, 平直度, 鼠标行为类型, 方向	各1个	6
总计		80

的大小是根据加入特征前后正域变化的大小来度量的。换句话说, 加入的特征能让越多的样本划入正域, 则这个特征的重要度越大。但是, 每次要加入一个新的特征时, 都需要计算所有未加入的特征相对于已选择的特征集的重要度, 其计算量偏大。为了减少特征选择的计算量, 本文使用特征选择领域较为常用的 ReliefF 算法进行特征重要度的预排序, 在进行基于邻域粗糙集的特征选择前, 提前算出各个特征的重要度, 这样加入新特征时只用一维正域计算, 提高了特征选择的效率。

1.3.1 基于 ReliefF 的特征重要度算法

ReliefF 算法^[17]运行效率高, 对数据类型没有限制, 因此得到了广泛应用。这里用 ReliefF 算法来计算特征的重要度, 该算法计算每个特征的重要度如下式所示。

$$W(A) = W(A) - \frac{\sum_{j=1}^k \text{diff}(A, R, H_j)}{mk} + \frac{\sum_{C \notin \text{class}(R)} \left[\frac{P(C)}{1 - P(\text{class}(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C)) \right]}{mk} \quad (4)$$

上式中, $W(A)$ 为特征 A 的重要度, m 为抽样次数, k 为最近邻样本个数, $P(C)$ 为原数据集中类别为 C 的样本所占比例, $M_j(C)$ 为类 $C \notin \text{class}(R)$ 中第 j 个最近邻样本, H_j 表示与 R 同类的样本集中最近的第 j 个样本, $\text{diff}(A, R, H_j)$ 表示样本 R 和 H_j 在特征 A 上的差, 如下式。

$$\text{diff}(A, R, H_j) = \frac{|R[A] - H_j[A]|}{\max(A) - \min(A)} \quad (5)$$

将得到的 $W(A)$ 按照逆序进行排序, 即可得到特征重要度序列 WsigOrder。基于 ReliefF 的特征重要度算法如下。

算法 1 基于 ReliefF 的特征重要度算法

输入 用户集 U , 抽样次数 m , 最近邻样本个数 k , 特征个数 n

输出 特征重要度序列 WsigOrder

1) 初始化特征权重 $W = \{0, \dots, 0\}$;

2) for $i = 1:m$

① 从用户集 U 中随机选择一个样本 R ;

② 从 R 的同类样本集中找到离 R 最近的 k 个

样本 H_j ($j=1, \dots, k$), 从每一个不同类样本集中找到离 R 最近的 k 个样本 $M_j(C)$ ($j=1, \dots, k$);

for $A = 1:n$

计算每个特征的重要度 $W(A)$;

end for

end for

3) WsigOrder = sort(W' , descend');

4) return WsigOrder

1.3.2 基于邻域粗糙集的特征选择算法

在根据 ReliefF 算法得到各个特征的重要度后, 轮流选择重要度最大的特征加入约简集合, 如果加入该特征后正域的样本增多, 就将此特征加入约简集合, 否则, 删除此特征。继续选择除约简集合外其他重要度最大的特征, 直到所有的样本都是正域的样本。此外, 在邻域粗糙集中, 邻域 δ 的设置是很重要的, 为了能够得到较好的效果, 本文选用文献[18]中使用的标准差方法来确定邻域的大小。基于邻域粗糙集的特征选择算法如下。

算法 2 基于邻域粗糙集的特征选择算法

输入 用户集 U , 鼠标行为特征集 C , 用户标识 D , 鼠标特征重要度序列 WsigOrder, 邻域 δ

输出 约简后特征集 red

1) 初始化 red = $\emptyset$, 正域样本 Pos = $\emptyset$, 待约简样本 sample = U , num = 0;

2) while sample $\neq \emptyset$

num = num + 1;

$a = \text{WsigOrder}(\text{num})$;

$DT = \langle \text{sample}, \text{red} \cup a, D \rangle$;

计算 DT 下的正域样本 Pos;

if Pos $\neq \emptyset$

red = red $\cup a$;

sample = sample - Pos;

else

break;

end if

end while

3) return red

1.4 信任模型

鼠标行为身份认证是一种基于行为的生物认证方法, 会因为情绪、压力等因素而表现得不稳

定。如果仅收集一个鼠标行为则难以正确判断出用户的身份,收集较多数量的鼠标行为进行身份认证,将导致认证时间较长,增加计算机系统的安全风险。为了减少身份认证时间,本文使用信任机制来持续性地判断用户的身份,其基本思想是系统对当前用户真实性的判断取决于该用户的行为与真实用户的行为的差异。如果真实用户与当前用户之间存在很大的偏差,则系统会降低对用户的信任度,反之会上升,如果系统对用户的信任度过低,则用户将被锁定在系统之外。非法用户的大多数鼠标行为是非法的,因此能够在较少的鼠标操作下被检测出来,进而减少身份认证时间。

1.4.1 信任函数

在信任模型中,信任值的变化量不是固定的,而是根据分类得分动态改变,通过信任函数衡量它的变化量。本文引入式(6)、式(7)所示的信任函数。

$$\Delta\text{Trust}(cs_i) = \min \left\{ -D + D \left(\frac{1+C}{C + \exp\left(-\frac{cs_i - A}{B}\right)} \right), C \right\} \quad (6)$$

$$\text{Trust}_i = \min \left\{ \max \left\{ \text{Trust}_{i-1} + \Delta\text{Trust}(cs_i), 0 \right\}, 100 \right\} \quad (7)$$

上式中, cs_i 表示第 i 个鼠标行为被分类器判断为合法行为的概率, Trust_i 表示收集第 i 次鼠标行为后的信任值, A 表示奖励和惩罚之间的阈值, B 表示 sigmoid 函数的宽度, C 和 D 表示奖惩上限。

本文通过大量实验得到较优的信任函数的参数,得到 $A=0.5$, $B=0.04$, $C=1$, $D=1$ 。将信任函数进行可视化,可得信任值的变化与分类得分的关系如图 2 所示,信任函数的曲线是对称的,在分类得分的值接近于 1 时,信任值的变化量为 1;在分类得分的值接近于 0 时,信任值的变化量为 -1。

1.4.2 信任模型阈值

合理地设置信任模型的阈值可以让系统收集较少的鼠标行为数据就能正确地判断出非法用户,并且合法用户不会被系统错误地判断为非法用户。在 Balabit 鼠标数据集中,表现较差的用户

16 的信任值变化如图 3 所示;在 DFL 鼠标数据集中,表现较差的用户 1 的信任值变化如图 4 所示。通过分析信任值波动较大的用户数据,可以看出其信任值维持在 95 以上,因此可以选取信任模型的阈值为 95 进行相关的实验。

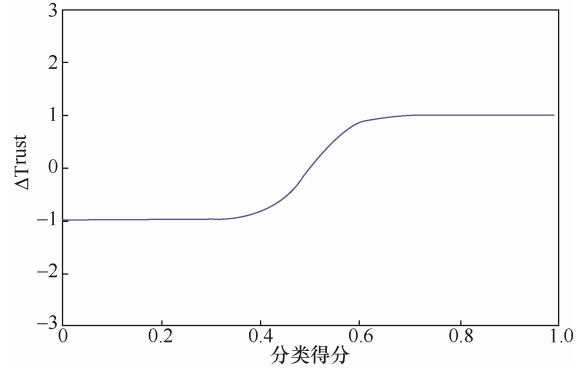


图 2 信任值的变化与分类得分的关系
Figure 2 Relationship between change of trust value and classifier score

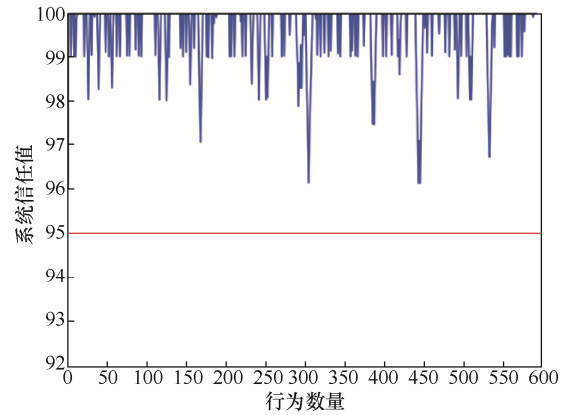


图 3 在 Balabit 数据集上用户 16 的信任值变化
Figure 3 Changes in the trust value of user 16 on the Balabit dataset

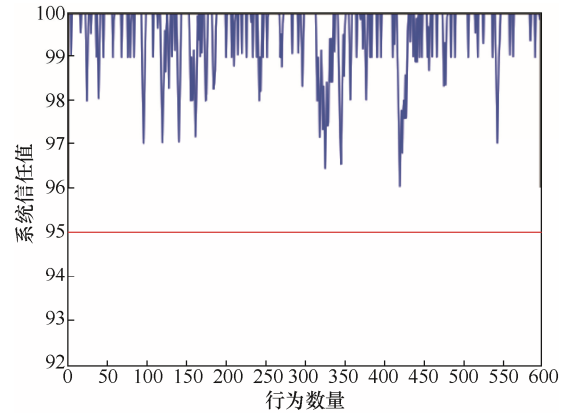


图 4 在 DFL 数据集上用户 1 的信任值变化
Figure 4 Changes in the trust value of user 1 on the DFL dataset

2 实验结果与分析

2.1 数据集

本文使用的数据集为 Balabit 和 DFL 公开鼠标数据集。Balabit 鼠标数据集来自 10 名连接到远程服务器的桌面客户端工作的用户，数据以会话为单位进行保存，包括时间戳、鼠标按钮的状态、鼠标指针的坐标。DFL 鼠标数据集收集于一个不受控制的环境下的 21 名用户，用户被要求在计算机上安装一个后台服务，以收集他们的鼠标数据，该数据集的存储格式和 Balabit 数据集一样。本文选择 DFL 鼠标数据集中会话数量最多的 10 名用户进行相关实验。

2.2 实验流程

实验首先对各个用户的鼠标行为进行划分，进而提取出相应的鼠标行为特征，然后使用 ReliefF 算法求得特征的重要度，并结合邻域粗糙集进行特征选择，得到约简后的鼠标行为特征。为了防止鼠标数据集不平衡而产生实验偏差，实验每轮确定 1 名用户作为合法用户，从该用户中随机选择 2 000 个鼠标行为作为正样本，从其他 9 名用户中共随机选择 2 000 个鼠标行为作为负样本，共 4 000 个鼠标行为数据作为数据集，并把该数据集按照 7:3 的比例分为训练集和测试集。使用二分类算法进行训练集的训练，得到该用户的身份认证模型，基于这个身份认证模型进行实验。每名用户的实验重复 10 次取结果的平均值，最后取这 10 名用户的实验结果的平均值作为最终的实验结果。

2.3 评价指标

本文选取 5 个评价指标对实验结果进行评估。

1) 准确率 (ACC): 表示预测结果与真实结果一致的比率。

$$ACC = \frac{\text{分类正确的样本}}{\text{总样本}}$$

2) 错误接受率 (FAR): 表示非法用户被错误识别为合法用户的概率。

$$FAR = \frac{\text{非法用户被识别为合法用户的样本}}{\text{非法用户的样本}}$$

3) 错误拒绝率 (FRR): 表示合法用户被错误识别为非法用户的概率。

$$FRR = \frac{\text{合法用户被识别为非法用户的样本}}{\text{合法用户的样本}}$$

4) ANIA (average number of imposter actions): 表示非法用户被检测之前可以操作的鼠标行为数量。

5) 特征维数: 表示特征空间的维度大小，影响分类模型的复杂度。

2.4 实验结果分析

在本文选择的鼠标行为特征的基础上，结合安全领域较为常用的支持向量机、随机森林和 XGBoost 作为分类模型进行分析，分别记为 Method1、Method2、Method3，同时与文献[19]、文献[20]、文献[21]中的实验方法进行对比分析。文献[19]和文献[20]提取了 39 个鼠标行为特征，在分类模型上分别使用的是随机森林和 K 近邻。文献[21]提取了 58 个鼠标行为特征，使用支持向量机作为分类模型。不同方法在 Balabit 数据集和 DFL 数据集上的实验结果见下文。

(1) 特征选择的结果

特征选择算法的结果比较如表 2 所示，从表中可以看出，改进后的特征选择算法相比之前，特征维数更少，准确率更高。与特征选择前相比，特征维数大幅度下降，身份认证性能几乎不变。本文结合 ReliefF 属性重要度和邻域粗糙集的特征选择算法能有效去除鼠标行为的无关或冗余特征。

表 2 特征选择算法的结果比较
Table 2 Comparison of feature selection results

特征选择算法	特征维数	准确率
无	80	99.02%
ReliefF-NRS	37	98.96%
ReliefF	46	96.54%
NRS	39	97.78%

(2) 固定鼠标行为数量的认证效果

收集固定数量的用户鼠标行为进行身份认证，并将本文提出的方法与其他文献的实验方法进行对比，根据前文的分析，选择的鼠标行为数量为 10。从图 5~图 7 的实验结果可知，在 Balabit 数据集上，Method1、Method2、Method3 的各个评价指标都优于前 3 种文献的实验方法；在 DFL 数据集上，Method1、Method2 的错误拒绝率略高于文

献[20], 其他评价指标均优于前 3 种文献的实验方法。实验效果最好的是 Method3, 即使用 XGBoost 作为分类算法, 在 Balabit 数据集上该方法的准确率为 98.96%, 错误接受率为 0.49%, 错误拒绝率为 1.58%; 在 DFL 数据集上该方法的准确率为 99.62, 错误接受率为 0.19%, 错误拒绝率为 0.53%。

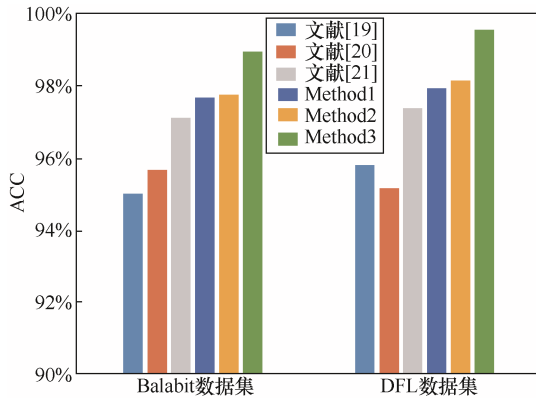


图 5 在 Balabit 和 DFL 数据集上不同方法的 ACC
Figure 5 ACC of different methods on Balabit and DFL datasets

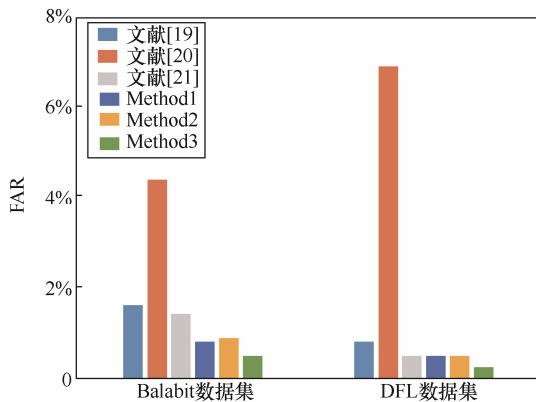


图 6 在 Balabit 和 DFL 数据集上不同方法的 FAR
Figure 6 FAR of different methods on Balabit and DFL datasets

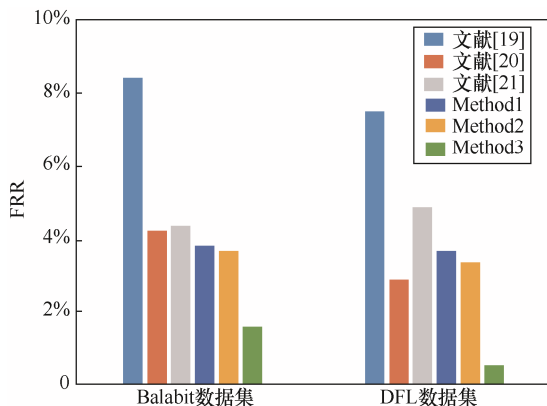


图 7 在 Balabit 和 DFL 数据集上不同方法的 FRR
Figure 7 FRR of different methods on Balabit and DFL datasets

从图 8 可以看出, 本文方法提取的鼠标行为特征维数量明显少于文献[21], 与文献[19]、文献[20]接近。综合比较可以看出, 本文方法具备一定的鲁棒性, 在两个数据集的各评价指标上都有稳定的认证结果, 优于其他文献的认证效果。

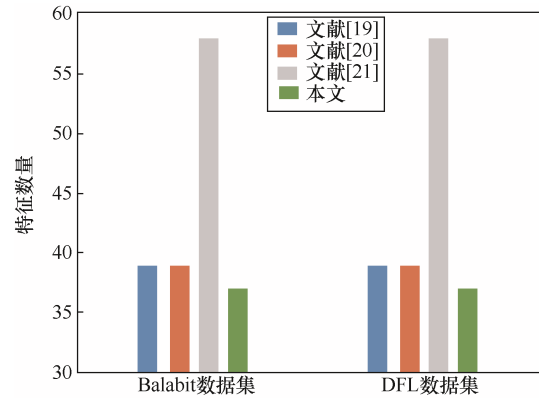


图 8 不同方法提取的特征数量
Figure 8 The number of behavior features extracted by different methods

(3) 结合信任模型的认证效果

在结合信任模型的实验中, 首先通过每次收集的鼠标行为进行分类得分的计算, 即分类器判断该鼠标行为为合法行为的概率, 然后信任函数基于此概率计算出信任值的变化量, 当信任值降低到阈值以下就会判定当前用户为非法用户, 并把信任值恢复为 100, 继续判断测试集中剩余鼠标行为的合法性, 从而得到平均收集多少个鼠标行为可以判断该用户为非法用户。从表 3 和表 4 可以看出, 其他文献中的实验方法结合信任模型表现最好的是文献[21], 在 Balabit 和 DFL 数据集上, 该方法的 ANIA 分别为 8.62 和 8.02; 本文所提方法认证效果最好的是 Method3, 在 Balabit 和 DFL 数据集上, Method3 的 ANIA 分别为 7.83 和 6.55。结合信任模型后, 本文提出的方法相对于其他文献的实验方法, 依然具有优势。

与前一个实验(固定鼠标行为数量的认证效果)相比, 固定 10 个鼠标行为, 最后各个方法得到的 ANIA 接近 10, 而从这个实验的结果中可以看出, 在 Balabit 数据集上, 每个实验方法结合信任模型都能降低 ANIA, 最多降低了 2.17; 在 DFL 数据集上, 除了文献[19]结合信任模型会略微提高 ANIA, 其他实验方法都会降低 ANIA, 最多降

低了 3.45。因此，结合信任模型能进一步减少非法用户在被系统识别出来之前能执行的鼠标行为数量。

表 3 在 Balabit 数据集上，不同方法结合信任函数的 ANIA
Table 3 ANIA of different methods combining trust functions on Balabit datasets

实验方法	ANIA
文献[21]	9.96
文献[22]	8.94
文献[23]	8.62
Method1	9.69
Method2	9.03
Method3	7.83

表 4 在 DFL 数据集上，不同方法结合信任函数的 ANIA
Table 4 ANIA of different methods combining trust functions on DFL datasets

实验方法	ANIA
文献[21]	11.21
文献[22]	8.13
文献[23]	8.02
Method1	7.83
Method2	7.43
Method3	6.55

(4) 外部用户的认证效果

为了验证本文方法对于外部用户有一定的鲁棒性，在 DFL 数据集中使用未参加训练的其他 5 名用户进行身份认证，结果如表 5 所示。从表中可以看出，由于认证模型缺乏外部用户的鼠标行为信息，准确率有所下降，但还是能以较高的准确率成功识别出，所以本文方法对于外部用户也具有有良好的鲁棒性。

表 5 外部用户的认证准确率
Table 5 Authentication accuracy of external users

用户	准确率
2	95.09%
4	95.97%
5	97.67%
6	94.11%
7	97.26%

3 结束语

本文研究了基于鼠标行为的持续身份认证问题，首先，针对鼠标行为特征中存在冗余或无关

特征，结合了 ReliefF 属性重要度和邻域粗糙集进行特征选择，降低特征维数；其次，为了在收集更少的鼠标行为下正确判断出非法用户，结合了信任模型进行持续身份认证；最后，通过实验证明本文方法能在较少的鼠标行为数量和特征下，达到较好的实验效果。

人脸、指纹都可以被模仿，而鼠标行为是长期养成的行为习惯，较难被模仿，所以本文方法具有一定的防伪性。本文提出的身份认证方法可以作为传统身份认证方法的辅助手段，持续地监控计算机系统使用者的身份，提升系统的安全性。

参考文献：

- [1] 荆继武. 网络可信身份管理的现状与趋势[J]. 信息安全研究, 2016, 2(7): 666-668.
JING J W. The development status and tendency of internet trusted identity management[J]. Journal of Information Security Research, 2016, 2(7): 666-668.
- [2] JOSHI V B, RAVAL M S. Adaptive threshold for fingerprint recognition system based on threat level and system load[J]. Procedia Computer Science, 2020, 171: 498-507.
- [3] LIU Q, JIANG B, ZHANG J L, et al. Semi-supervised uncorrelated dictionary learning for colour face recognition[J]. IET Computer Vision, 2020, 14(3): 92-100.
- [4] KUMAR M R, ARTHI K. An effective non-cooperative iris recognition system using hierarchical collaborative representation-based classification[J]. The Journal of Supercomputing, 2020, 76(1): 1-14.
- [5] AHMED A A E, TRAORE I. Biometric recognition based on free-text keystroke dynamics[J]. IEEE Transactions on Cybernetics, 2014, 44(4): 458-472.
- [6] CHANG T Y, TSAI C J, YEH J Y, et al. New soft biometrics for limited resource in keystroke dynamics authentication[J]. Multimedia Tools and Applications, 2020, 79(31-32): 23295-23324.
- [7] AHMED A A E, TRAORE I. Detecting computer intrusions using behavioral biometrics[C]//3rd Annual Conference on Privacy, Security and Trust. 2005: 91-98.
- [8] AHMED A A E, TRAORE I. A new biometric technology based on mouse dynamics[J]. IEEE Transactions on Dependable & Secure Computing, 2007, 4(3): 165-179.
- [9] ZHENG N, PALOSKI A, WANG H. An efficient user verification system via mouse movements[C]//Proceedings of the 18th ACM Conference on Computer and Communications Security. 2011: 139-150.
- [10] ZHENG N, PALOSKI A, WANG H. An efficient user verification system using angle-based mouse movement biometrics[J]. ACM Transactions on Information and System Security, 2016, 18(3): 1-27.
- [11] 徐剑, 李明洁, 周福才, 等. 基于用户鼠标行为的身份认证方法[J]. 计算机科学, 2016, 43(2): 148-154.
XU J, LI M J, ZHOU F C, et al. Identity authentication method

- based on user's mouse behavior [J]. Computer Science, 2016, 43(2): 148-154.
- [12] YILDIRIM M, ANARIM E. Session-based user authentication via mouse dynamics[C]//27th Signal Processing and Communications Applications Conference. 2019: 1-4.
- [13] 田杰, 宋礼鹏. 基于鼠标及窗口行为的持续身份认证研究[J]. 中北大学学报(自然科学版), 2020, 41(6): 511-519, 526.
TIAN J, SONG L P. Research on continuous identity authentication based on mouse and window behavior[J]. Journal of North China University(Nature Science Edition), 2020, 41(6): 511-519, 526.
- [14] CHONG P, TAN Y X M, GUARNIZO J, et al. Mouse authentication without the temporal aspect-what does a 2D-CNN learn[C]//IEEE Security and Privacy Workshops (SPW). 2018: 15-21.
- [15] 沈超, 蔡忠闽, 管晓宏, 等. 基于鼠标行为特征的用户身份认证与监控[J]. 通信学报, 2010, 31(7): 68-75.
SHEN C, CAI Z M, GUAN X H, et al. User authentication and monitoring based on mouse behavioral features[J]. Journal on Communications, 2010, 31(7): 68-75.
- [16] KONONENKO I. Estimating attributes: analysis and extensions of relief[C]//European Conference on Machine Learning. 1994: 171-182.
- [17] LADHA L, DEEPA T. Feature selection methods and algorithms[J]. International Journal of Advanced Trends in Computer Science and Engineering, 2011, 3(5): 1787-1797.
- [18] 奖志军, 易华蓉. 一种基于图像金字塔光流的特征跟踪方法[J]. 武汉大学学报(信息科学版), 2007, 32(8): 680-683.
JANG Z J, YI H R. A feature tracking method based on image pyramid optical flow[J]. Journal of Wuhan University(Information Science Edition), 2007, 32(8): 680-683.
- [19] ANTAL M, EGYED-ZSIGMOND E. Intrusion detection using mouse dynamics [J]. IET Biometrics, 2019, 8(5): 285-294.
- [20] ALMALKI S, CHATTERJEE P, ROY K. Continuous authentication using mouse clickstream data analysis[C]//International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage. 2019: 76-85.
- [21] YILDIRIM M, ANARIM E. Novel feature extraction methods for authentication via mouse dynamics with semi-supervised learning [C]//Innovations in Intelligent Systems and Applications Conference (ASYU). 2019: 1-6.

[作者简介]



易聪 (1995-), 男, 四川达州人, 重庆邮电大学研究生, 主要研究方向为智能信息处理和数据挖掘。



胡军 (1977-), 男, 湖北监利人, 重庆邮电大学教授、博士生导师, 主要研究方向为粒计算、粗糙集、智能信息处理和数据挖掘。