L-009

マウスカーソルの軌跡を用いた個人認証

Personal authentication using mouse cursor path

岸 海渡　　堂薗 浩
Kaito Kishi　Hiroshi Dozono

1. はじめに

個人認証とは，システム利用者が登録者本人であるかを識別するプロセスで，情報化社会が加速する現代で重要な要素となる．特にパソコンの利用者が増加し機密情報を取り扱う機会が多くなり，その個人認証にはパスワード入力，指紋認証，顔認証が代表例として挙げられる．パスワード入力はキーボード入力による盗み見の危険性があることに加え，個人情報として入手しやすいものを使用していることからセキュリティ上の脆弱性，指紋，顔認証は高精度なものの専用の機器の必要性が課題となる．そこでマウスカーソルの軌跡からログイン認証時の画面上の座標分布，速度変化等の特徴量を利用し機械学習による個人の識別を行う個人認証システムの作成，精度の検討を行った．

2. 実験方法

本実験で作成した座標取得プログラムは，ログイン時の認証を想定して

1. 右クリックでデータの取得を開始
2. 3 秒間の経過時間，X 座標，Y 座標を配列に逐次記録
3. 3 秒経過後，配列データを npy ファイルで保存

の 3 工程が実行される．加えてスリープ状態から起動する際の再現として，座標取得プログラム実行と同時に全画面のウィンドウを疑似的なスリープ画面として表示させる．

2. 1　実験 1　座標データを用いた分類

はじめに，座標データのみを用いた分類を行う，作成したプログラムを被験者 5 名を対象に配布，説明を行い 1 人 10 回の軌跡データのサンプルを取得した．

2. 2　実験 2　座標データと速度データでの分類

次に，座標データと速度データを特徴量とした分類を行う．被験者がよりスムーズにサンプルデータの取得が可能となるように，かつ実際の認証に近づけるため，全画面ウィンドウを Windows のデスクトップスクリーンショットにして疑似性を高めた．(図 1)さらに，のぞき見された場合の対策を想定し，画面の録画を行う．その後録画データを本人以外の被験者が視聴し，カーソルの動きを真似して追加のデータを 2 データずつ取得し，本人データ 10 データ，他者の模倣データ 8 データをサンプルとして使用する．また，本人データと模倣データを合わせた際の異常検出を行う．

図 2 に実験 2 で取得した被験者のカーソル軌跡の例を示す．取得したデータを分類用に調整を行い 1 列のデータに再配列する．1 列のデータに並べられた 3 秒間の X 座標，Y 座標，速度データを特徴量として，教師データ，検証データに分割して機械学習の手法である多層パーセプトロン(MLP)[1]やランダムフォレスト[2]を用いて分類の精度を 5 回分の平均を用いて確認する．異常検出には学習手法のアイソレーションフォレスト[3]を用いる．

異常検出では各パラメータの組み合わせごとに 10 回平均を求める．各ユーザの認証率が最も高くなるパラメータを選別し，その 5 パターンの組み合わせで 5 人の自他分類を行った際の確率を求め，検証する．

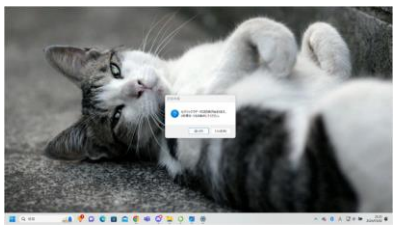

図 1:プログラム実行時のスクリーンショット

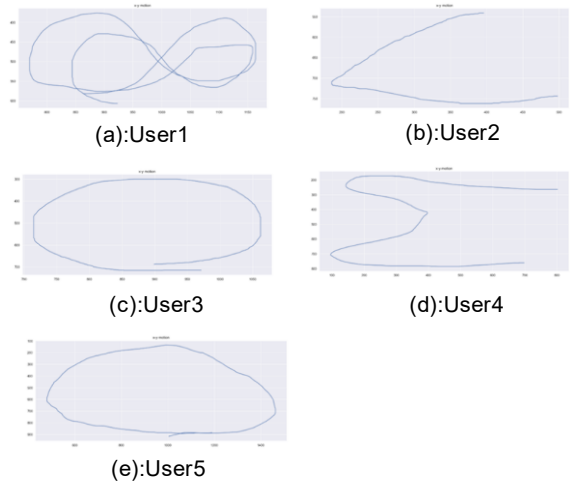

図 2:5 人の被験者軌跡データ

3. 実験結果と検討

実験では座標データのみを用いた分類，速度データを加えた分類，異常検出の 3 つを行った．座標データのみの実験 1 の分類では MLP を用いて全体の正答率が 94.6%となり，高い数値が記録できたものの個別認証率が 80%のものがあったため，パラメータの調整を行い，適したパラメータ 条件を調べた．最も認証率が高かった結果を 表 1. (a) に示す．

次に速度データを追加した実験 2 での MLP を用いた 5 分類でパラメータを変更して最も高い認証率の結果を表 1.(b.1)に示す．被験者 5 名のうち 4 名分のデータは 80%以上の認証率だったが，座標データのみを用いた分類に比べ数値が低い結果となった．加えて，User5 の認証率は 66.66%となり，すべての被験者で高い精度で分類できたとは言えない結果となった．そこで，異なる機械学習手法であるランダムフォレストを用いて再分類を行う．ランダムフォレストは決定木とアンサンブル学習を組み合わせた予測，分類に長けたアルゴリズムで，データ数が多いほど分類精度の向上を望むことができる．総要素数が座標のみに比べ多くなった本実験ではそのメリットが合致したため採用した．

ランダムフォレストを用いて分類を行った結果を 表 1. (b.2)に示す. MLP を用いたときに比べ 個別の認証精度は高く, 最も低い正答率である user5 でも 95%と高い認証率を示すことができた.

最後に被験者 1 人ごとの異常検出を行う. 他人の軌跡を模倣したデータを各 2 データずつ取得し, 本人データ 10, 他人データ 8 で, 本人データ 80%を学習データとして異常検知を行った.

評価のために調整するパラメータは n_estimators(決定木の本数決定), contamination(モデル感度)の 2 つである. n_estimators は利用する決定木の本数で, この数値が高いほど精度は上がるが, 高すぎると過学習の要因となる. contamination はデータセット内の異常の割合を変えることでモデルの感度を調整するパラメータで, 異常の割合が高いほど, モデルは異常を検知しやすくなるが、正常データを誤って異常と判定する可能性も増える.

各ユーザで本人認証率と他人拒否率が最も高いパラメータの組み合わせと確率を表 2(a)に示す. 最良パラメータの決定基準は, 本人認証率と他人拒否率がともに 80%以上となることとした. この 5 つの組み合わせで 5 人の異常検出を行った結果の中で最も数値が高かったものを表 2(b) に示す. User2 から User4 は, このパラメータで本人認証率, 他人拒否率がともに 80%以上と高い精度での異常検知ができていたが, User5 は他人認証率が 26.25%, User1 に関しては異常検知の数値は高かったものの, 本人認証率が 50%を下回るという結果になった. 表 2(a)での User1 が最良のパラメータの数値は他の結果に比べ決定木の数値が高く, かつ 90%以上の本人認証率を出していないことから, 一筆書きのような図形を描く軌跡に比べ, 再現性の低い軌跡の識別は難しいことがわかる.

表 1. (a):座標データの 5 分類

ユーザ名	正答率[%]
user1	100
user2	100
user3	85
user4	96
user5	100
全体	94.6

表 1. (b.1):座標+速度データ(MLP)

ユーザ名	正答率[%]
user1	88
user2	89.32
user3	95
user4	95
user5	66.33
全体	85.32

表 1. (b.2):座標+速度データ(ランダムフォレスト)

ユーザ名	正答率[%]
user1	100
user2	100
user3	100
user4	100
user5	95
全体	98.66

表 2. (a):各ユーザの最も評価の良いパラメータと確率

ユーザ名	n_estimators	contamination	本人認証率[%]	本人拒否率[%]	他人拒否率[%]	他人認証率[%]
user1	100	0.1	80.00	20.00	95.00	5.00
user2	110	0.2	100.00	0.00	98.75	1.25
user3	40	0.2	100.00	0.00	97.50	2.50
user4	40	0.2	100.00	0.00	96.25	3.75
user5	50	0.4	90.00	10.00	90.00	10.00

表 2. (b):user3 最良パラメータの結果

ユーザ名	本人認証率[%]	本人拒否率[%]	他人拒否率[%]	他人認証率[%]
user1	40.00	60.00	96.25	3.75
user2	100.00	0.00	90.00	10.00
user3	85.00	15.00	98.75	1.25
user4	95.00	5.00	96.25	3.75
user5	100.00	0.00	73.75	26.25

4. まとめ

本研究ではログイン認証時のマウスカーソルの軌跡を用いた個人認証プログラムの作成を行い, 座標データを用いた分類と, 速度データを追加した際の分類と異常検出の検証と考察を行い, 実現性を示した. 座標データのみの 2 値分類において本人受入率はどれも 90%を上回ったものの, 5 分類の個別の正答率が 90%を下回る場合があった. 次に速度データを特徴量に追加して 5 分類検証を行った結果, 座標データのみの場合に比べて個人の正答率, 全体の正答率ともに低い値を示した. そこで, 分類器をランダムフォレストに変更して検証を行ったところ, 全体では 98.66%, 個別の分類においても 95%以上の正答率をどのサンプルにおいても示した. その後, 本人データと異常データを識別した際の異常検出の精度を調べた. 全サンプルにおいて, 個別に設定したパラメータにおいては高い数値を出したものの, 共通して適用するために 1 つのパラメータで検証したところ, 異常検知の感度が高く, 本人も認証されない場合があった.

今後の課題として, 異常検知の精度について本人を異常として拒否してしまう確率をさらに減らす手法として, アイソレーションフォレスト以外の分類方法の活用や分類の閾値の調整が考えられる.

謝辞

本研究を行うにあたって, 日頃から親切かつ丁寧なご指導をいただいた堂薗 浩准教授並びに諸先生方に厚く御礼申し上げます. また, 本研究についてご助言, ご協力いただいた本研究室の学部生, 大学院生の皆様に深く感謝申し上げます.

参考文献

[1] Technical Note, ”多層パーセプトロン”, https://hkawabata. github. io/technical-note/note/ML/mlp. html869
[2] IBM, ” ランダム・フォレストとは” , https://www. ibm. com/jp-ja/topics/random-forest.
[3] 株式会社システムインテグレータ, ” 異常検出アルゴリズムの代表格「Isolation Forest」とは?”, https://products. sint. co. jp/aisia-ad/blog/what-is-isolation-forest