タッチベース認証における公正な誤認証率評価フレームワークの提案

工藤 雅士† 高橋 翼‡ 牛山 翔二郎† 山名 早人§

† 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1
‡ LINE 株式会社 〒160-0004 東京都新宿区四谷一丁目 6 番 1 号 四谷タワー23 階
§ 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: †§{kudoma34, s-ushiyama, yamana}@yama.info.waseda.ac.jp, ‡tsubasa.takahashi@linecorp.com

あらまし スマートフォンの第三者利用を防止するため，近年，タッチベース認証が注目されている．タッチベース認証は，スマートフォンのログイン認証が破られた際の砦になり得るものであり，ユーザのタッチ操作を継続的に監視することにより，ユーザの操作性を損なわずに不正な使用を検出することができる．スマートフォンの普及以降，タッチベース認証に関する数多くの研究が行われているが，標準的な性能評価指針が存在しないため，異なるタッチベース認証手法間での性能比較が困難となっている．そこで本研究では，従来の性能評価手法の問題を明らかにすると共に，タッチベース認証における公正な評価手法として「誤認証率評価フレームワーク」を提案する．提案するフレームワークでは，１）実運用を想定したデータセットの構築，２）新たな誤認証率評価指標の導入を行うことで，異なる認証手法間での公正で現実的な誤認証率の比較を可能にする．

キーワード 生体認証，スマートフォン，HCI，データセット，タッチストローク，性能評価

1. はじめに

近年爆発的に普及したスマートフォンには，個人情報や機密データを保護するために，ログイン時に本人認証を行う機能が導入されている．標準的な本人認証として，パターンやパスコードといったユーザの記憶に基づいた記憶認証と，顔認証[1]や指紋認証[2]といったユーザの身体的特徴（または行動的特徴）を利用した生体認証が挙げられる．記憶認証はユーザが自由に設定でき，変更が容易である等の利点がある一方で，忘却や覗き見によるログイン情報の盗難[3]，画面の汚れによる推測[4]などのリスクが存在する．生体認証は記憶認証と比較して，ログイン情報の記憶や認証のための操作などユーザへの負担が少なく，また模倣や複製のリスクが低く安全性が高いとされてきた．しかし，近年の技術進歩により高解像度の画像または動画から生体情報を複製することにより，認証の突破が可能であることが報告されている[5]．こうした現状から，新たな認証方式の需要が高まりを見せている．

新たな認証方式としては，タッチベース認証が注目を集めている．タッチベース認証は，「スマートフォンはタッチスクリーン上で操作が行われる」という点に着目し，スマートフォン利用者のタッチデータを継続的に取得し認証に利用する手法である．タッチベース認証では，タッチ速度やタッチ座標，タッチ圧力などのストローク特性に基づいて，現在のユーザが正規ユーザか非正規ユーザかを判定する．タッチベース認証は，認証のために追加の操作を必要とせず，またログイン時の標準的な認証と組み合わせて使用できるため，スマートフォンの操作性を損なうことなくセキュリティ性を高めることができるという利点がある．

タッチベース認証はスマートフォンが普及し始めた 2010 年代初頭から現在までに多くの手法が提案されてきたが，その実用化は未だに進んでいない．実用化が進んでいない要因の一つとして，タッチベース認証の性能の優劣が判定しにくい点が挙げられる．具体的には，性能評価の方法に指針が存在せず，手法間の比較が困難となっている現状がある．

そこで本稿では，タッチベース認証の実用化に向けた試みとして，タッチベース認証の公正な誤認証率評価フレームワークを提案する．本フレームワークに基づいてタッチベース認証の性能評価を実施することにより，現実の状況に即した条件下で認証手法間の性能比較が可能になる．また，ユーザが持つストローク特性の多様性を考慮した性能比較が可能になる．

本稿では次の構成をとる．2 節でタッチベース認証に関する予備知識を説明し，3 節でタッチベース認証の関連研究を述べる．続いて，4 節で提案手法の概要を説明し，5 節で評価実験の設計方法，6 節で評価実験および結果について説明する．7 節でタッチベース認証の誤認証率評価フレームワークについて検討を行い，8 節で本稿をまとめる．

2. 予備知識

2.1 タッチベース認証

タッチベース認証は，(1) 本人登録，(2) 継続認証，(3) 分類器更新の 3 つのフェーズで構成される．タッチベース認証の全体像を図 1 に示す．各フェーズにお

いて実施される処理内容は以下の通りである．なお，以下では，本人を「正規ユーザ」，本人以外を「非正規ユーザ」と表現する．

1. 本人登録
   - スマートフォンのセンサーからタッチデータ取得（データ取得）
   - 外れ値の除去や正規化等の前処理（データ前処理）
   - 前処理後のタッチデータから，継続認証で使用する特徴量を抽出（特徴量抽出）
   - 正規ユーザか非正規ユーザかを判定する分類器生成（分類器生成）
2. 継続認証
   - 新たに取得したタッチデータが正規ユーザのものであるかを分類器で評価し，正規ユーザとの類似度もしくは本人確率を出力（データ分類）
   - 分類器から出力されたスコアを基に，正規ユーザであるかを判定（意思決定）
3. 分類器更新
   - 正規ユーザのタッチデータは一定ではなく，時間と共に変化する．この変化に対応するために，一定の間隔で分類器を更新（データ適応）

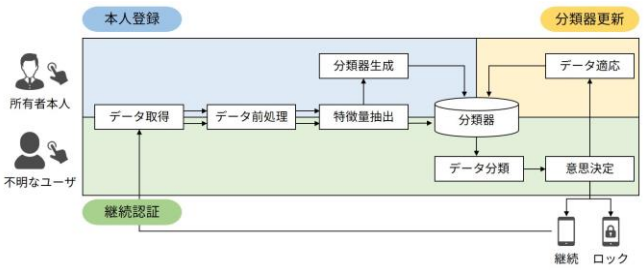

図 1　タッチベース認証の全体像

2.2 評価指標

タッチベース認証は，正規ユーザを正例（Positive），非正規ユーザを負例（Negative）として，二値分類を行う．そのため，認証システムの予測結果は，表 1 の混同行列に示される 4 つのクラスに分類できる．

表 1　混同行列

	Predicted Positive	Predicted Negative
Actually Positive	TP (True Positive)	FN (False Negative)
Actually Negative	FP (False Positive)	TN (True Negative)

タッチベース認証のような継続認証の性能は，主にセキュリティ性と操作性の二つの観点から評価が行われる．セキュリティ性は非正規ユーザの侵入を防ぐ性能を評価する指標であり，個人情報を扱うスマートフォンでは高い水準が求められる．一方で，非正規ユーザの判定を厳しくした場合，認証システムが正規ユーザを誤判定してしまう確率も増加する．正規ユーザの誤判定が頻繁に発生する場合，ユーザはその都度操作を中断しなければならなくなり，操作性が低下する．タッチベース認証では，このようなトレードオフの関係を持つセキュリティ性と操作性を評価する指標として，FAR (False Acceptance Rate)，FRR (False Rejection Rate), EER (Equal Error Rate)が使用される．FAR, FRR, EER の詳細は以下の通りである．

1. FAR (False Acceptance Rate)

FAR は非正規ユーザを正規ユーザとして誤判定する確率を表す．FAR はその値が低いほど非正規ユーザの検出率が高く，セキュリティ性が高いことを示す．FAR は以下の式(1)で算出される．

$$FAR = \frac{FP}{FP + TN} \cdots\cdots (1)$$

2. FRR (False Rejection Rate)

FRR は正規ユーザを非正規ユーザとして誤判定する確率を表す．FRR はその値が低いほど正規ユーザを拒否する頻度が低く，操作性が高いことを示す．FRR は以下の式(2)で算出される．

$$FRR = \frac{FN}{TP + FN} \cdots\cdots (2)$$

3. EER (Equal Error Rate)

EER は，FAR と FRR が等しい場合におけるユーザの誤判定確率を示し，以下の式(3)で算出される．

$$EER = \frac{FAR + FRR}{2} \text{（但し，FAR=FRR の時）} \cdots\cdots (3)$$

EER はセキュリティ性を示す FAR と，操作性を表す FRR に基づいて算出されるため，継続認証の評価においてセキュリティ性と操作性のトレードオフを測定するための評価指標として使用される．

3. 関連研究

3.1 公正な評価に向けた取り組み

タッチベース認証の公正な評価に向けた取り組みとして，タッチデータセットの公開が挙げられる．スマートフォンはモデルの移り変わりが激しく，またユーザによって持ち方や操作方法が異なるため，タッチデータを取得する際の条件を全体で統一するのは困難である．したがって，タッチデータセットの公開はタッチベース認証の公正な評価を行う上で重要取り組みとなる．タッチベース認証に関するオープンデータセットを表 2 にまとめる．なお，本稿執筆時点[1]では Frank データセット[6]と Antal データセット[8]の 2 つが入手可能である．

[1] 2022 年 12 月 26 日 12 時時点

Frank データセット[6]は，41 人の実験参加者から Android 操作時の垂直方向および水平方向のスワイプデータを取得したデータセットであり，2013 年に公開された．実験参加者は 4 つの Android（Nexus One, Nexus S, Galaxy S, Droid Incredible）のうち 1 つを操作して，テキストを読むタスクと横並びに配置された 2 つの画像の違いを探すタスクをそれぞれ実施した．

Antal データセット[8]は，71 人の実験参加者から Android 操作時の垂直方向および水平方向のスワイプデータを取得したデータセットであり，2015 年に公開された．実験参加者は 8 つの Android（機種の記載はなし）のうち 1 つを操作して，画面に表示されたテキストを読むタスクと，画像のギャラリーを閲覧するタスクをそれぞれ実施した．前者のタスクでは垂直方向のスワイプ，後者のタスクでは水平方向のスワイプがそれぞれ取得された．

3.2 タッチベース認証の性能

タッチベース認証では，分類モデルとして，典型的な機械学習モデルである k-NN と SVM が広く採用されている．タッチベース認証における k-NN と SVM の性能を表 3 および表 4 にまとめる．

2.2 項で説明したように，FAR と FRR はトレードオフの関係を持ち，認証システムが採用する閾値に応じて変化するため，その値だけでは手法間での優劣の判定は困難である．EER については，タッチベース認証の性能評価時に実施された交差検証の平均値を示しており，テストデータセットを構成するユーザによって値が変化する可能性がある．そのため，EER も同様にその値だけでは手法間での性能の比較は困難である．

また，タッチベース認証では訓練とテストで同一の負例ユーザを使用して性能評価を行う場合が多い（表 3，表 4）．しかし，タッチベース認証を実際に使用する場面では，継続認証時に現れる負例ユーザを事前に特定することは困難である．したがって，現在用いられている既知の負例ユーザを用いた性能評価は，実際の使用状況を想定した公正な評価であるとは言えない．

4. 提案概要

4.1 提案の目的

タッチベース認証では性能評価の方法に指針が存在しないため，手法毎に独自で性能評価が実施されている．また，評価の際に訓練とテストで同じ負例ユーザを使用する場合が多く（表 3，表 4），実際の使用状況を反映した評価が行われていない．さらにストローク特性はユーザによって異なるため，訓練とテストで使用される負例ユーザによって誤認証率に差が生じる．

本稿では，タッチベース認証において公正かつ現実的な評価を実施するためのフレームワークを提案することにより，各認証手法間での性能比較を容易にし，タッチベース認証の実用化の推進を目指す．

4.2 貢献

本稿の貢献を以下にまとめる．

- 負例ユーザ選択方法に関する問題提起：テスト時に使用する負例ユーザを，訓練時と同じユーザに設定した場合と違うユーザに設定した場合で EER を算出し，その差について統計的検定を実施して有意な差が生じることを示す．
- 小さいデータセットでの評価に関する問題提起：タッチベース認証の評価結果が訓練およびテストで選択される負例ユーザに依存して変化することを示す．
- 公正なタッチベース認証フレームワーク提案：現実的な状況を再現して公正なタッチベース認証の性能評価を行うためのフレームワークを提案し，タッチベース認証の実用化を推進させる．
- タッチベース認証の頑健性評価指標の提案：ユーザの組み合わせによる分類性能のブレを反映させた新たな評価指標を提案する．

5. 実験設計

5.1 分類モデルの構築

本稿では，最新のタッチベース認証手法である Zaidi らの手法[14]に基づいて分類モデルを構築する．Zaidi ら[14]は，継続認証に使用する分類器を訓練データに基づいて動的に複数選択し，アンサンブルで本人判定を行う手法を提案した．Zaidi ら[14]は，動的アンサンブル選択（DES: Dynamic Ensemble Selection）の手法を 12 種類検証し，Randomised Reference Classifier（以下，DES-RRC）[18]が様々な認証シナリオで一貫して優れた性能を発揮することを確認した．そこで本稿では，タッチベース認証で一般的に使用されている機械学習分類器である k-NN と SVM に加えて，DES-RRC を分類モデルとして採用する．また，DES-RRC で選択される分類器の候補は Zaidi らの手法に基づいて，k-NN, SVM, 決定木, ナイーブベイズ, ロジスティック回帰, ニューラルネットワークを採用し，分類モデルは 10 ストローク分の判定結果の平均値をもとに本人判定を行うものとした．各分類器を訓練する際に実施したパラメータサーチの概要を表 5 に示す．

表 2　タッチベース認証のオープンデータセット

年	データセット	取得ユーザ数	特徴量の種類数	入手可否（執筆時点）[1]
2013	Frank [6]	41	30	○[2]
2013	Serwadda [7]	190	28	×[3]
2015	Antal [8]	71	15	○[4]
2016	Mahbub [9]	48	24	×[5]
2019	Syed [10]	31	18	×[6]

表 3　タッチベース認証の性能（k-NN）

年	著者	データセット	テスト用負例ユーザ	FAR (%)	FRR (%)	EER (%)
2013	Frank et al. [6]	Frank [6]	既知	-	-	0.00 – 4.00
2013	Serwadda et al. [7]	Serwadda [7]	未知	-	-	14.00
2015	Shen et al. [11]	Private	既知	0.1	28.52	-
2018	Kumar et al. [12]	Private	既知	14.02	3.87	-
2021	Incel et al. [13]	Private	既知	0.30	5.85	5.39
2022	Zaidi et al. [14]	Frank [6]	既知	-	-	0.94
		Serwadda [7]	既知	-	-	1.39
		Antal [8]	既知	-	-	2.75
		Mahbub [9]	既知	-	-	27.84

表 4　タッチベース認証の性能（SVM）

年	著者	データセット	テスト用負例ユーザ	FAR (%)	FRR (%)	EER (%)
2013	Frank et al. [6]	Frank [6]	既知	-	-	0.00 – 4.00
2013	Serwadda et al. [7]	Serwadda [7]	未知	-	-	13.10
2015	Shen et al. [11]	Private	既知	0.10	20.42	-
2016	Kumar et al. [12]	Private	既知	18.12	3.04	-
2018	Meng et al. [15]	Private	既知	4.95	4.37	-
2018	Chang et al. [16]	Frank [6]	既知	-	-	0.76
2018	Fierrez et al. [17]	Frank [6]	既知	-	-	5.30
		Serwadda [7]	既知	-	-	4.40
		Antal [8]	既知	-	-	4.40
		Mahbub [9]	既知	-	-	10.90
2019	Syed et al. [10]	Syed [10]	既知	-	-	18.70
2021	Incel et al. [13]	Private	既知	0.04	3.88	3.50
2022	Zaidi et al. [14]	Frank [6]	既知	-	-	4.51
		Serwadda [7]	既知	-	-	4.04
		Antal [8]	既知	-	-	5.40
		Mahbub [9]	既知	-	-	35.90

表 5　各分類モデルのパラメータ候補一覧（Zaidi らの手法[14]に基づいて設定）

分類モデル	パラメータ	値
k-NN	Number of neighbors	[1,2,3,4,5,6,7,8,9,10]
	Algorithm	KD tree
	Distance metric used for finding neighbors	Euclidian
SVM	Regularization parameter, C	[0.001,0.01,0.1,1,10,25,50,100,1000]
	Kernel	RBF
	Kernel coefficient, γ	1/number of features
	Tolerance for stopping criterion	$1e^{-3}$
NB	-	-
DT	Maximum depth of the tree	[none,5,10,15,20,30]
	Minimum number of samples to split	[2,4,6,8,10]
	Minimum number of samples at a leaf Number	[1,2,3,4,5]
LR	-	-
MLP	Number of hidden layers	1
	Number of hidden nodes	50

[2] http://www.mariofrank.net/touchalytics/
[3] http://www2.latech.edu/%20phoha/BTAS-2013.htm
[4] https://www.ms.sapientia.ro/~manyi/bioident.html
[5] https://umdaa02.github.io/
[6] http://zasyed.com/jss18dataset.html

5.2 認証シナリオ

本稿では，現在も入手可能であり，タッチベース認証の評価において広く使用されている Frank データセット[6]と Antal データセット[8]を使用して評価実験を実施する．Frank データセットは垂直方向と水平方向のストロークを含むデータセットであり，データ収集の際に設けられた休憩時間を区切りとして，1 ユーザ毎に最大 7 セッションで構成される．Antal データセットは，垂直方向と水平方向のストロークを含むデータセットであるが，データ上に区切りは存在しない．

本稿では，先行研究[6][14][16][17]で広く採用されている認証シナリオに基づいて，ストローク方向（垂直・水平）と，Frank データセットの場合はセッション（セッション内・セッション間）を考慮した 6 つの認証シナリオ（表 6）を設定し，評価実験を実施する．

表 6　認証シナリオ

表記	データセット	シナリオ
$FRK_{v\text{-}intra}$	Frank [6]	セッション内垂直
$FRK_{h\text{-}intra}$	Frank [6]	セッション内水平
$FRK_{v\text{-}inter}$	Frank [6]	セッション間垂直
$FRK_{h\text{-}inter}$	Frank [6]	セッション間水平
ANT_v	Antal [8]	垂直
ANT_h	Antal [8]	水平

5.3 データセットの構築

本稿では，先行研究[6][14][17]で広く採用されているデータセット構築方法を採用した．以下にデータセット構築の手順を示し，図 2 に全体の流れをまとめる．

1. 訓練・テストデータセットの構築（正例）

   認証システムにおいて正例として認識するユーザ（以下，正例ユーザ）を，全ユーザの中から 1 ユーザ選出する．正例ユーザのタッチデータから時系列順に，最も古い連続した 40 ストロークを抽出し，正例の訓練データセットD_{train_pos}を構築する．D_{train_pos}で使用していない正例ユーザのタッチデータ全て抽出し，正例のテストデータセットD_{test_pos}を構築する．

2. 訓練データセットの構築（負例）

   認証システムの訓練時に負例として認識するユーザ（以下，訓練用負例ユーザ）を，正例ユーザ以外からランダムに 4 ユーザ選出する．各訓練用負例ユーザのタッチデータから，ランダムに 10 ストロークずつタッチデータを抽出し，負例の訓練データセットD_{train_neg}を構築する．

3. テストデータセットの構築（負例）

   認証システムの評価時に負例として認識するユーザ（以下，テスト用負例ユーザ）を，正例ユーザ以外からランダムに 4 ユーザ選出する．各テスト用負例ユーザのタッチデータからランダムに同数のストロークを抽出し，D_{test_pos}と同数の負例テストデータセットD_{test_neg}を構築する．

本稿では，データセット構築の都合上，60 ストローク以上のデータを持つユーザのみを対象に評価を行う．また，手順 3 において選出されるテスト用負例ユーザを，手順 2 で選出される訓練用負例ユーザと同一のユーザに設定した場合を既知データセットと定義し，手順 2 で選出される訓練用負例ユーザと異なるユーザに設定した場合を未知データセットと定義する．そして，ストローク特性はユーザによって異なるため，各データセットの評価結果は，評価対象のユーザを正規ユーザとして選出する全てのパターンで EER を算出し，その平均値で評価する．つまり，評価対象のユーザが全体で 20 人いた場合，20 通りのデータセットを EER で評価し，その平均値を報告する．

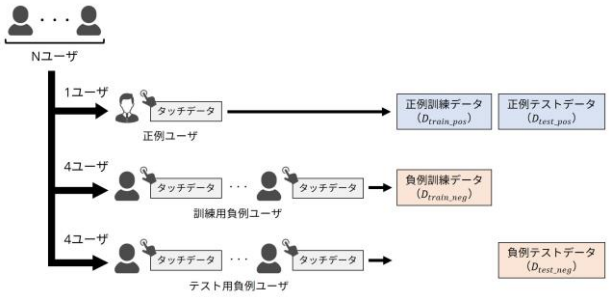

図 2　評価用データセットの構築方法

6. 評価実験

6.1 既知および未知データセット間の EER 比較

本項では，5.3 項で構築した既知および未知データセットの EER の違いを確認する．各分類器および各認証シナリオにおける評価結果を図 3 に示す．図 3 より，テスト時の負例ユーザが未知の場合は，既知の場合と比較して EER が有意に高くなることを確認した．実際にタッチベース認証が使用される環境では負例ユーザは未知であるため，EER が高いとしても未知データセットで性能を評価することが望ましいといえる．

6.2 テスト用負例ユーザによる EER のばらつき検証

タッチベース認証の評価実験では，1 名の正例ユーザに対して，ランダムに複数の負例ユーザを選出して評価用データセットを構築する．ユーザによってストローク特性は異なるため，同じ正規ユーザに関する評価であっても，テスト時に選出される負例ユーザによって EER に差が生じる可能性が考えられる．本項では，未知データセット構築時に選出されるテスト用負例ユーザによって，EER に有意な差が生じるかを検証する．検証にあたり，テスト用負例ユーザの選出方法を変えた 2 つのデータセットを構築して EER の比較を行う．

1. 正例ユーザとの判別が困難な 4 ユーザ（上位 4 ユーザ）をテスト用負例ユーザに選出
2. 正例ユーザとの判別が容易な 4 ユーザ（下位 4 ユーザ）をテスト用負例ユーザに選出

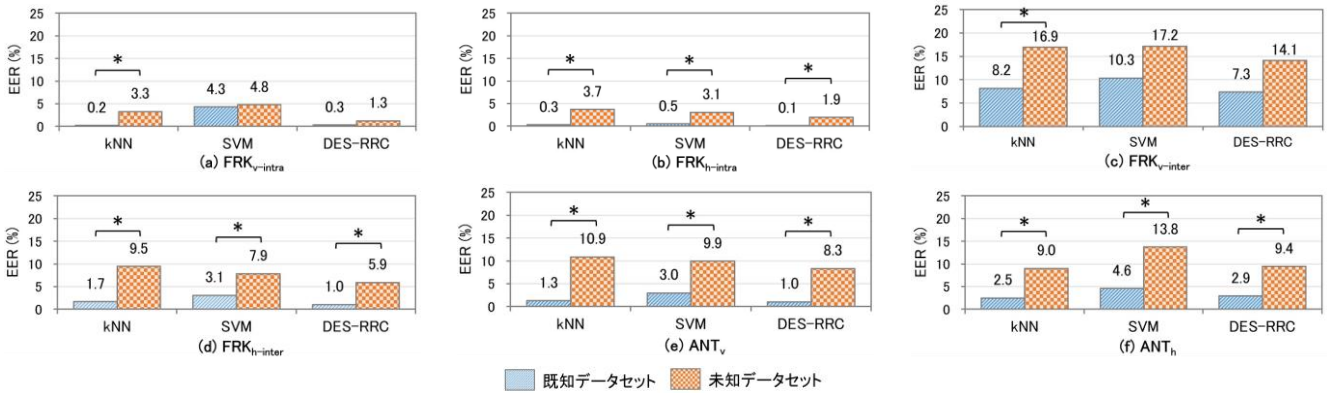

図 3　既知データセットと未知のデータセットの EER 比較結果（*：p＜0.05）

データセット 1 およびデータセット 2 において，正例ユーザとの判別の困難さは，以下の手順で算出される EER に基づいて設定する．

1.　正例・負例ユーザの設定

正例ユーザを，全ユーザの中から 1 ユーザ選出する．その後，テスト用負例ユーザを，正例ユーザを除くユーザから 1 ユーザ選出する．正例ユーザとテスト用負例ユーザ以外のユーザを，訓練用負例ユーザとして設定する．

2.　訓練データセットの構築

正例ユーザのタッチデータから時系列順に，最も古い連続した 40 ストローク分のタッチデータを抽出し，正例の訓練データセットD_{train_pos}を構築する．その後，各訓練用負例ユーザからランダムに同数のストロークを抽出し，全体で 40 ストローク分のタッチデータを含む負例の訓練データセットD_{train_neg}を構築する．

3.　テストデータセットの構築

D_{train_pos}で使用していない正例ユーザのタッチデータ全て抽出し，正例のテストデータセットD_{test_pos}を構築する．負例テストデータセットD_{test_neg}は，テスト用負例ユーザのタッチデータからD_{test_pos}と同数のストロークをランダムに抽出して構築する．

4.　EER の算出

D_{train_pos}とD_{train_neg}で分類器を訓練し，D_{test_pos}とD_{test_neg}で正例ユーザに対するテスト用負例ユーザの EER を算出する．その後，テスト用負例ユーザと訓練用負例ユーザを変えて，負例ユーザ全パターンにおける EER を算出し，ユーザの順位付けを行う．

手順 1 から手順 3 に示されるデータセット構築の流れを図 4 にまとめる．各分類器および各認証シナリオにおける評価結果を図 5 に示す．図 5 より，負例として選出されるユーザによって，最大で EER 10%以上の有意な差が生じることを確認した．

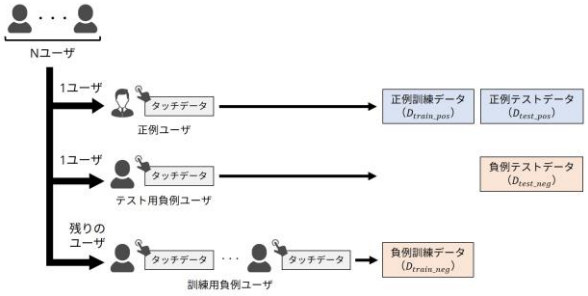

図 4　ユーザ間の判別困難性評価用データセットの構築方法

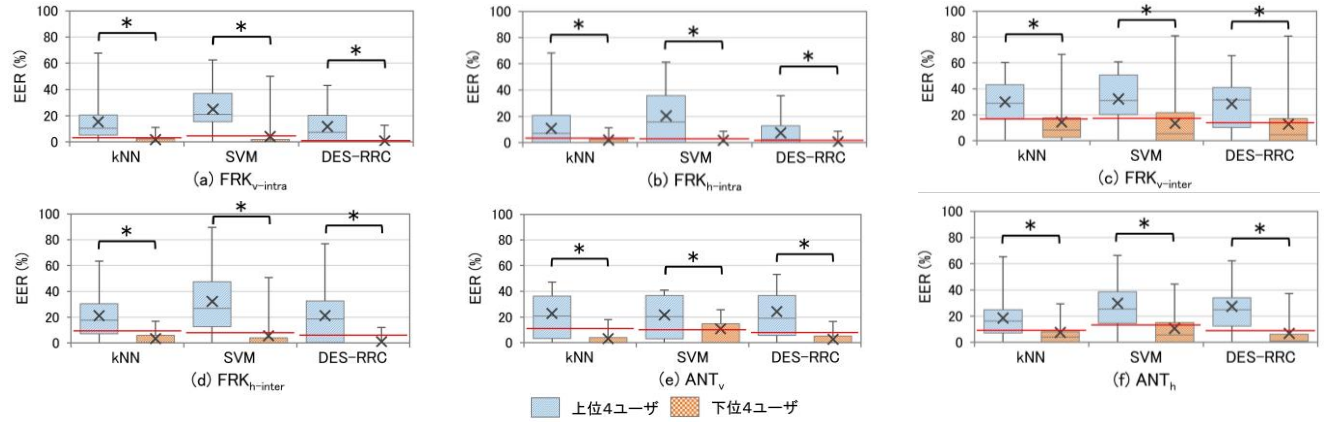

図 5　正例ユーザとの判定が困難な上位 4 ユーザと下位 4 ユーザの EER 比較結果
（*：p < 0.05，赤線は 6.1 項の図 3 における未知データセットの評価結果を表す）

7. 誤認証率評価フレームワークの提案

7.1 公正な評価データセットの構築

タッチベース認証実用化のためには，現実的な評価データセットを構築し，複数の観点から構成に評価実験を行う必要がある．6 節で得られた評価結果を基に，タッチベース認証における評価データセット構築のポイントを以下にまとめる．

- 継続認証時に現れる負例ユーザを事前に特定することは困難であるため，訓練時とテスト時では異なる負例ユーザを使用して評価する．
- 継続認証の評価結果は，訓練とテストで選択される負例ユーザに依存する．そのため,負例ユーザをランダムに選択した場合の EER に加えて，使用データセットの中で最も分類が困難な場合と最も分類が容易な場合の EER も併せて報告する．

7.2 新たな評価指標の提案

$FRK_{h\text{-}intra}$ シナリオにおいて，6.2 項で算出した正例ユーザ・負例ユーザの全組み合わせにおける EER を図 6 に示す．図 6 に示されるように，多くの組み合わせで EER が 0%となっている一方で，一部ユーザの組合わせでは EER が 50%を超え，著しく分類性能が低下していることが確認できる．EER でのみ分類性能を比較すると，こうした分類が困難なユーザの組み合わせが存在することを正しく理解することは難しい．そのため，新たな評価指標を検討する．

図 6 で示した $FRK_{h\text{-}intra}$ シナリオにおける EER と，その累積相対度数の関係を図 7 に示す．図 7 において，AUC（Area Under the Curve: 曲線下面積）が大きいほど分類性能が高いことを示す．図 7 における各分類器の EER，AUC，標準偏差および EER を超える領域の AUC を表 7 にまとめる．EER と AUC を比較すると DES-RRC が最も優れているが，図 6 に示されるように DES-RRC においても一部分類が困難なユーザの組み合わせが存在する．これを理解した上で性能比較をするために，「EER を超える領域の AUC」を新たな評価指標として検討する．標準偏差のような，性能が良い方向と悪い方向の双方のブレではなく，本指標を用いて性能が悪い方向へのブレに焦点を当てて AUC を評価することにより，分類が困難なユーザの存在を踏まえた性能評価が可能になる．実際に，本指標を用いて各分類器の性能を比較すると，k-NN と RRC が同等の性能であることが理解できる．

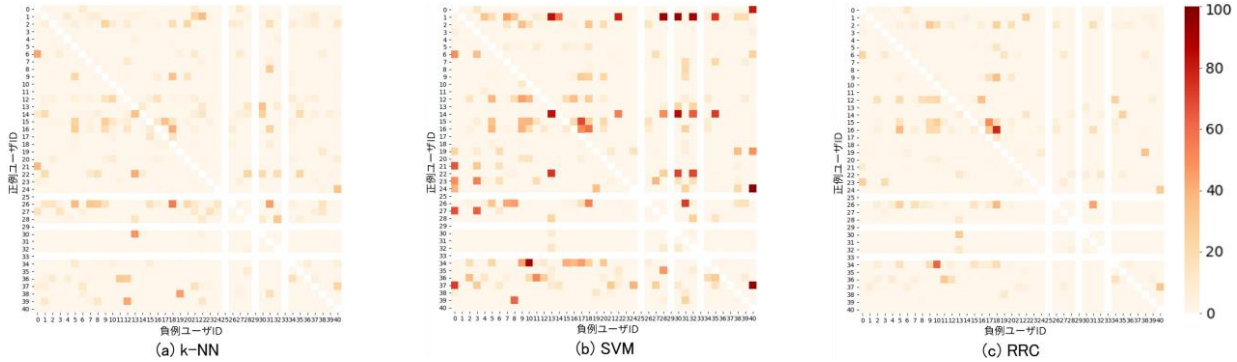

図 6　$FRK_{h\text{-}intra}$ シナリオにおける正例ユーザ（行）と負例ユーザ（列）の EER 行列（色が濃いほど EER が高いことを示す）

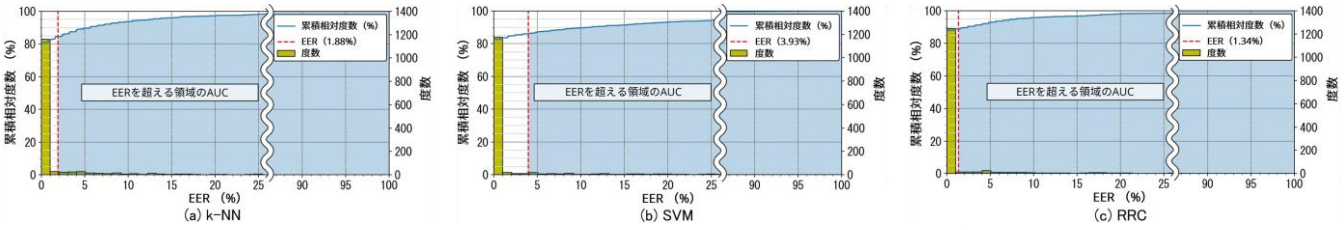

図 7　$FRK_{h\text{-}intra}$ シナリオにおける EER と累積相対度数の関係

表 7　$FRK_{h\text{-}intra}$ シナリオにおける EER，AUC，標準偏差および EER を超える領域の AUC

分類器モデル	EER (%)	AUC	標準偏差	EER を超える領域の AUC
k-NN	1.88	0.981	6.02	0.881
SVM	3.93	0.961	12.92	0.724
DES-RRC	1.34	0.987	5.47	0.882

8. まとめ

本稿では，タッチベース認証の実用化に向けた新たな試みとして，タッチベース認証の公正かつ現実的な評価を実施するためのフレームワークについて検討および提案を行った．テスト時に使用する負例ユーザが既知の場合と未知の場合で EER を比較した評価実験では，前者の EER が有意に低くなることを確認した．また，テスト時に選出する負例ユーザを判別が困難なユーザに設定した場合と判別が容易なユーザに設定した場合で EER を比較した評価実験では，最大で EER 10%以上の有意差が生じることを確認した．これらの結果を踏まえ，公正かつ現実的な評価を行うために，１）訓練時とテスト時では異なる負例ユーザを使用する，２）負例ユーザをランダムに選択した場合の EER に加えて，最も分類が困難な場合と最も分類が容易な場合の EER も併せて報告する，以上の２点をデータセット構築のポイントとしてまとめた．

また，評価実験の際に一部のユーザの組み合わせで EER が 50%を超え，判別が困難なパターンがあること確認した．このような判別が困難なパターンの存在は，EER だけでは正しく理解ができないため，EER と AUC を組み合わせた新たな評価指標の提案を行った．

今後の課題としては，本稿で提案した評価指標の詳細な検証や，本フレームワークが他の分類器や認証手法に適用可能か検証を実施することなどが挙げられる．

謝　辞

この研究は 2022 年度国立情報学研究所 CRIS 共同研究の助成を受けています。

参　考　文　献

[1] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain, "Continuous Authentication of Mobile User: Fusion of Face Image and Inertial Measurement Unit Data", Proc. of the IEEE Int. Conf. on Biometrics (ICB'15), pp. 135-142, 2015.

[2] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Handbook of Fingerprint Recognition", Springer, 2009.

[3] F. Tari, A. A. Ozok, and S. H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords", Proc. of the 2nd Symposium on Usable Priv. and Secur. (SOUPS'06), pp. 56-66, 2006.

[4] A. Aviv, K. Gibson, and E. Mossop, "Smudge Attacks on Smartphone TouchScreens", Proc. of the 4th USENIX Conf. Offensive Technol., pp. 1-7, 2010.

[5] I. Echizen, and T. Ogane, "BiometricJammer: Method to Prevent Acquisition of 40 Biometric Information by Surreptitious Photography on Fingerprints", IEICE Trans. Inf. Syst., Vol. E101D, No.1, pp. 2-12, 2018.

[6] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication", IEEE Trans. Inf. Forensics and Secur., pp. 136-148, 2013.

[7] A. Serwadda, V. V. Phoha, and Z. Wang, "Which Verifiers Work?: A Benchmark Evaluation of Touch-based Authentication Algorithms", IEEE Biometrics Theory, Applications and Syst. (BTAS'13), 2013.

[8] M. Antal, Z. Bokor, and Z. Laszlo, "Information revealed from scrolling interactions on mobile devices", Pattern Recognition Lett., Vol. 56, pp. 7-13, Elsevier, 2015.

[9] Y. Mahbub, S. Sakar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results", IEEE Biometrics Theory, Applications and Syst. (BTAS'16), 2016.

[10] Z. Syed, J. Helmick, S. Banerjee, and B. Cukic, "Touch gesture-based authentication on mobile devices: The effects of user posture, device size, configuration, and inter-session variability", Journal of Syst. and Software, Vol. 149, Elsevier, pp. 158-173, 2019.

[11] C. Shen, Y. Zhang, X.Guan, and R.A. Maxion, "Performance analysis of touch-interaction behavior for active smartphone authentication", IEEE Trans. Inf. Forensics Secur. Vol. 11 (3), pp. 498-513, 2015.

[12] R. Kumar, P.P. Kundu, and V.V. Phoha, "Continuous authentication using one-class classifiers and their fusion", Proc. of the IEEE 4th Int. Conf. on Identity, Secur., and Behavior Analysis (ISBA'18), 2018, Vol. 2018-Janua. IEEE, pp. 1-8, 2018.

[13] O.D. Incel, S. Gunay, Y. Akan, Y. Barlas, O.E. Basar, G.I. Alptekin, and M. Isbilen, "DAKOTA: Sensor and Touch Screen-Based Continuous Authentication on a Mobile Banking Application", IEEE Access, Vol. 9, pp. 38943-38960, 2021.

[14] A. Z. Zaidi, C. Y. Chong, R. Parthiban, and A. S. Sadiq, "A framework of dynamic selection method for user classification in touch-based continuous mobile device authentication", Journal of Inf. Secur. and Applications, Vol. 67, 103217, 2022.

[15] W. Meng, W. Li, and D.S. Wong, "Enhancing touch behavioral authentication via cost-based intelligent mechanism on smartphones", Springer, 2018.

[16] I. Chang, C.Y. Low, S. Choi, and A. Beng Jin Teoh, "Kernel deep regression network for touch-stroke dynamics authentication", IEEE Signal Process. Lett., Vol. 25 (7), 2018.

[17] J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales, "Benchmarking Touchscreen Biometrics for Mobile Authentication", IEEE Trans. Inf. Forensics Secur., Vol. 13 (11), pp. 2720-2733, 2018.

[18] T. Woloszynski and M. Kurzynski, "A probabilistic model of classifier competence for dynamic ensemble selection", Pattern Recognition, Vol. 44 (10-11), Elsevier, pp. 2656-2668, 2011.