

コントローラ操作の入力密度に着目した対戦型格闘ゲームにおけるプレイヤー認証

川本 岳史^{1,a)} Maharage Nisansala Sevwandi Perera¹ Franziska Zimmer¹ 小林 良輔¹
Mhd Irvan¹ 山口 利恵¹

概要：e スポーツの公平性を担保する上でプレイヤーの認証は不可欠である。試合中のプレイヤーの行動バイオメトリクスを用いることで認証性能が向上することが報告されているが、これには行動情報以外に試合の状況を示す付随情報が必要であった。本研究ではコントローラ操作の入力密度に着目し、プレイ中のデータから重要な部分を特定して認証に利用する手法を提案する。従来手法と同様の評価を行ったところ、提案手法は付隨情報を用いずとも従来手法に匹敵する識別性能を持つことを示した。さらに 60 名のプレイヤーによる実際のプレイデータを用いた評価において、提案手法は PR-AUC 94.2%，EER 0.52% という高い識別性能を達成した。本研究の成果は行動情報さえあればプレイヤーの認証を行える可能性を示唆するものである。

キーワード：行動バイオメトリクス、プレイヤー認証、e スポーツ、入力密度

Player Authentication in a Fighting Game Using the Input Density of Controller Operations

TAKESHI KAWAMOTO^{1,a)} MAHARAGE NISANSALA SEWANDI PERERA¹ FRANZISKA ZIMMER¹
RYOSUKE KOBAYASHI¹ MHD IRVAN¹ RIE SHIGETOMI YAMAGUCHI¹

Abstract: Ensuring a player's identity during a match is crucial for maintaining fairness in esports. While prior research reported that performance can be improved by extracting in-match data, this process required supplementary game-state information in addition to raw player behaviors. To address this limitation, we utilize the density of controller operations as a key to identify important segments from gameplay data for authentication purposes. The proposed method demonstrates performance comparable to prior work, despite not relying on external information. Furthermore, on a dataset of 60 players, our method achieved a PR-AUC of 94.2 % and an EER of 0.52 %. These results highlight the possibility of player authentication in esports using solely behavioral data, thereby eliminating the need for game-state information.

Keywords: behavioral biometrics, player authentication, esports, input density

1. はじめに

e スポーツは、ゲーム対戦をスポーツ競技として捉える呼称である。大会数や賞金額は年々増加しており、伝統的

なスポーツと同等の公平性が求められる。しかし e スポーツの大会や試合にはオンラインで行われるものも多く、公平性の担保における障害となっている。

e スポーツの公平性の脅威としてなりすましが挙げられる。これは名義上のプレイヤーになり変わってプレイする行為である。名義上のプレイヤーと実際のプレイヤーの一一致を確認する手段には、ゲームアカウントの所持をもって

¹ 東京大学
The University of Tokyo

a) kawamoto@yamagula.ic.i.u-tokyo.ac.jp

本人とする方法や、ゲームをプレイしている姿をカメラで映させる方法がある。しかしこれらは2人のプレイヤーが協力状態にある場合に共有や改竄が容易な認証情報を用いており、対策として十分でない。そこで、共有も改竄も難しい認証情報として行動バイオメトリクスを用い継続的に操作者を認証する手法が求められている。

ゲームにおける行動バイオメトリクス研究は、試合中などゲームのルールに従って行動する状況に注目してきた。例えば試合中のキャラクターの位置 [1] やプレイヤーの操作 [2,3] を用いる研究が挙げられる。さらに [4] は試合中と休憩中という異なるシーンのデータを用いたプレイヤー認証の精度を比較し、試合中のデータを抜き出すことで高い精度のプレイヤー認証が可能なことを実験的に確認した。しかし、このアプローチはデータを抜き出す際に試合の開始・終了時刻の一覧を追加で必要とし、プレイヤーの行動記録単体では動作しない。加えて実験的なデータセットでの検証であり実際の環境への適応可能性が未検証であった。

そこで本研究では入力密度、すなわちコントローラが操作される頻度に着目してシーン（試合中・休憩中）を推定し試合中のデータを抽出する。これにより、試合の開始・終了時刻に頼ることなくコントローラ操作単体で認証が行えることになる。先行研究との比較実験では、シーン推定に入力密度を用いることで PR-AUC（適合率 - 再現率曲線下面積）がわずかに低下、EER（等価エラー率）がわずかに向上する結果を示した。これはコントローラ操作のみで従来手法に匹敵する識別性能を持つことを意味している。さらにプレイヤー 60 名、307 試合分のデータセットを構築し、シーンごとの有効性、判定の集約方法、パラメタごとの影響を調査した。その中で、入力密度の高いデータのみを用いて試合単位でプレイヤーを識別することで PR-AUC 94.22 %、EER 0.52 % という高い識別性能を達成した。

本研究の貢献は以下の通りである。

- コントローラ操作の入力密度のみを用いて重要なシーンを抽出する手法を提案し、従来手法（外部情報あり）と同等の識別性能を外部情報なしに達成した。
- プレイヤー 60 名、307 試合分のデータセットを構築し、多様で実用に近い環境における提案手法の網羅的な評価を行った。

2. 関連研究

本研究はゲーム内のプレイヤーの行動情報を用いたプレイヤー認証に関する研究分野に位置付けられる。本節ではまず関連する研究アプローチを整理し、次に我々の先行研究とその課題を述べることで、本研究が取り組むべき点を明らかにする。

2.1 ゲームの行動情報を用いたプレイヤー認証

ゲームをプレイする際のプレイヤーの行動情報は、プレイヤーによる操作と、それを処理した結果であるゲーム内の情報に分けることができる。プレイヤーの操作を用いるものとして、対戦型格闘ゲームにおける試合中のコントローラ操作による認証 [2]、Minecraft におけるマウス操作による認証 [3]、VR ゲームにおけるコントローラ操作による認証 [5] が挙げられる。また、ゲーム内で処理された情報を用いるものとして、シューティングゲームにおけるプレイヤーの座標・所持品による認証 [1]、ロールプレイントゲームにおけるアイテム購買履歴による異常検知 [6]、マルチプレイヤーオンラインバトルアリーナにおけるプレイヤーの体力などのステータスを用いた異常検知 [7] などが挙げられる。

これらは主にゲームプレイ中の行動を用いたものであり、ゲームプレイの前後や待機時間中の行動を用いたものは限定的である。そこで我々の先行研究 [4] では、対戦型格闘ゲームのシーン（試合中・休憩中）ごとの行動とプレイヤー識別性能の関係を調査した。結果、試合中のデータのみを用いることで高い識別精度を得られることを示し、その理由として試合中データがコントローラ操作の頻度が高く、プレイヤーごとの個�性が強く現れていることを指摘した。しかし、以下が課題として残されていた。

- 試合中データを抽出するために試合の開始・終了時刻の一覧を必要とし、コントローラ操作の記録単体では動作しない。
- 試合中・休憩中データの融合がデータレベル [8] に留まり、試合中データの優位性が明らかでない。
- データセットが 9 名のプレイヤーが单一キャラクターでプレイした際のものであり、多様なプレイヤー・プレイ環境を反映した評価を行っていない。

2.2 本研究の位置付け

プレイヤーやキャラクターの行動情報を用いた認証は幅広く研究されている。これらはいずれもゲームプレイ中の行動に着目しており、我々はこれを支持する結果として、我々は試合中の行動情報を抽出することで認証性能を高められる可能性を確認した。

そこで本研究は以下に取り組む。第一に、試合時刻のような付随情報を必要としない認証を実現するため、コントローラ操作からシーン（試合中・休憩中）かを自動で判定し重要なデータを抽出する方法を提案する。第二に、試合中データの優位性を確認するため、異なるシーンのデータを決定レベルで融合する方法を検討し、性能を比較する。第三に、提案手法の一般性を示すため、大人数のプレイヤーが異なるキャラクターを使うような環境で記録されたデータセットを用いて評価を行う。

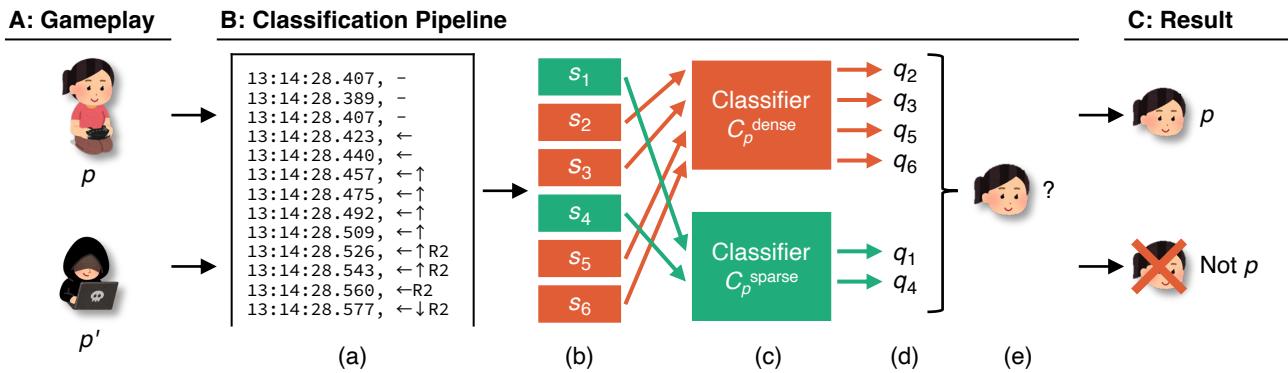


図 1 提案手法の全体像. プレイヤーがゲームをプレイし (A), そのコントローラ操作を記録 (B-a), 分割・分類 (B-b) する. その後分類器に入力され (B-c), 得られた結果 (B-d) を集約して (B-e), プレイヤーが p であるかどうかを判定する (C).

Fig. 1 The proposed method. A: Player plays the game; B-a: Record the controller operation logs; B-b: Segment and classify the logs; B-c: Input the segments to the classifiers; B-d: Calculate the probabilities; B-e: Aggregate the results; C: Determine whether the player is p .

3. 手法

本研究の目的は, コントローラの操作データからシーンを判定して重要な部分を抽出し, プレイヤーを認証することである. そのため, 既存手法 [4] を改良する.

3.1 問題設定と全体像

本研究ではゲームをプレイしている人物が誰であるかを識別することでプレイヤー認証を行う. すなわち, ゲームプレイ時の行動データとプレイヤーのリスト $\mathcal{P} = \{p_1, p_2, \dots\}$ およびプレイしていると主張するプレイヤー $p \in \mathcal{P}$ が与えられる. この時に, そのデータが p によるものか別人 p' によるものかを判定することに取り組む. 本研究では先行研究と同様に, $p' \in \mathcal{P} \setminus \{p\}$ として p 以外のプレイヤーを想定する.

これを実現する提案手法の全体像を図 1 に示す. これは先行研究に以下の変更を加えたものである.

- (1) セグメント分類 (B-b) : 試合状態に応じた分類から, 入力密度に基づく密・疎の分類に変更する.
- (2) 分類器の構成 (B-c) : 全セグメント単一の分類器から, 密・疎ごとに専用の分類器を用いるように変更する.
- (3) 意思決定の粒度 (B-e) : セグメント単位での決定から, それらを統合したバトル単位での決定に変更する.

3.2 ストリートファイター 6

本研究で扱うゲームは 1 対 1 対戦型格闘ゲームの『ストリートファイター 6』(SF6)^{*1} である. プレイヤーはコントローラを用いてキャラクターを操作する. 対戦相手の体力を先にゼロにした方がその“ラウンド”を勝利し, 先

に 2 ラウンドを勝利した方がその“バトル”の勝者となる.

本手法で用いるデータは SF6 プレイ時のコントローラ操作の履歴である. コントローラに計測装置を装着し, 1 フレーム (1/60 秒) ごとに押下されているボタンを記録する. 入力には 4 種類の矢印 (\leftarrow , \uparrow , \rightarrow , \downarrow), 9 種類のボタン (\circlearrowleft , \times , \triangle , \square , L1, L2, R1, R2, option) の計 13 種類が含まれ, 何も押下されていない場合はハイフン (-) が出力される.

3.3 前処理

データを一定の長さ θ フレームごとに区切りセグメントに分割していく. ラウンドの開始・終了タイミングによらず, 単純に θ フレーム分のデータを切り出す. 例えはあるバトルが 90 秒 (5400 フレーム), セグメント長が $\theta = 500$ の場合, 11 個のセグメントが得られる.

こうして区切られた各セグメントを密 (dense) または疎 (sparse) に分類する. セグメントに含まれるフレームの数を N_{all} , 入力を伴うフレームの数を N_{active} として入力密度を $N_{\text{active}}/N_{\text{all}}$ により求める. これが閾値 ρ を超えていれば密, 超えていなければ疎とする. その後, 行先研究と同様にそれぞれのセグメントに対して時間特徴量を計算し, その統計量を特徴量とする.

3.4 モデル分割

3.1 の設定に従い one-vs-rest 二値分類器を用いる. これは対象プレイヤーのデータを陽性, それ以外の全てのデータを陰性とする分類器である. 各プレイヤー $p \in \mathcal{P}$ に対して, 密なセグメントのみを用いる C_p^{dense} と疎なセグメントのみを用いる C_p^{sparse} を用意する. C_p^{dense} と C_p^{sparse} は対応するセグメントのみを用いて, 入力されたコントロー

^{*1} <https://www.streetfighter.com/6/>

ラの操作が p によるものか、他のプレイヤー $p' \in \mathcal{P} \setminus \{p\}$ によるものかを判定するよう訓練される。先行研究の結果から、ベースモデルには Random Forest を用いる。

3.5 アンサンブル

セグメントごとの差異による性能影響を防ぐため、本研究ではバトル単位の意思決定を採用する。いくつかのセグメントから成るバトル全体のコントローラ操作 $\mathcal{B} = \{s_1, s_2, \dots\}$ が与えられる。ここで以下のいずれかのアンサンブル手法を用いて、各セグメントを分類器に入力して得られたセグメント単位の判定をバトル全体で集約する。

3.5.1 多数決

まず各セグメント $s \in \mathcal{B}$ を密ならば C_p^{dense} 、疎ならば C_p^{sparse} に入力し、そのプレイヤーである確率 $q_p(s)$ を得る。その後確率のバトル全体での平均 $\bar{q} = \frac{1}{|\mathcal{B}|} \sum_{s \in \mathcal{B}} q_p(s)$ を求め、 $\bar{q} > 0.5$ ならばこのバトルは p のもの、 $\bar{q} \leq 0.5$ ならば p 以外のものだと決定する。

3.5.2 スタッキング

各セグメント $s \in \mathcal{B}$ をその疎密によらず $C_p^{\text{dense}}, C_p^{\text{sparse}}$ の両方に入力し、それぞれの確率 $q_p^{\text{dense}}(s), q_p^{\text{sparse}}(s)$ を得る。その後、これらのバトル全体での平均、標準偏差、最大値、最小値を求める。これらを特徴量とし p によるものかを判定する二値分類器 C_p^{fusion} を用いて、最終的な決定を行う。入力が単純なため、この分類にはロジスティック回帰を用いる。

4. 実験

本研究では既存手法を改良して試合状況の不要なプレイヤー認証手法を提案した。そこで、既存手法の改良による性能変化の確認と、提案手法の複数条件における性能比較を目的として実験を行う。

4.1 データセット

本研究では 2 種類のデータセットを使用する。小規模データセットは先行研究 [4] と同じものであり、実験環境下で収集された。9 名のプレイヤーの 99 バトル分のデータから成り、すべて同じキャラクターを使用している。そのため、プレイヤーの行動記録に加えてラウンドの開始・終了タイミングがフレーム単位で記録されている。

大規模データセットは新たに記録したデータセットであり、実際の大会において収集されたものである。60 名のプレイヤーの 307 バトル分のコントローラ操作から成り、各プレイヤーが使用するキャラクター、コントローラは異なる。そのため小規模データセットよりも多様で現実の状況を反映したデータセットだが、各ラウンドの開始・終了タイミングは記録されていない。

なお、これらのデータセットはいずれも東京大学大学院情報理工学系研究科倫理委員会の承認を得て収集された。

4.2 実験設計

本研究の提案手法の先行研究に対する差分は以下の 3 点である。

- (1) 状況の入力密度による疎密分類による代替
- (2) 密・疎それぞれ専用の分類器への振り分け
- (3) セグメント単位の決定の集約

これらを踏まえて、以下の 3 種類の実験を行う。

- 疎密分類の有効性の検証：(1) に対応
- 疎密分類の効果の検証：(2) および (3) に対応
- パラメタによる性能変化の確認

4.2.1 疎密分類の有効性の検証

各セグメントのシーンを入力密度による疎密分類で代替できるかを確認する。従来手法はセグメントの分類に各ラウンドの開始・終了時刻を用いている一方で、提案手法はセグメントの入力密度を用いている。そこで以下の 3 種類を比較することで、ラウンドの時刻という追加情報を用いずとも性能を維持できるかを確認する。

Active 従来手法の試合中データと提案手法の密なデータ

Idle 従来手法の休憩中データと提案手法の疎なデータ

All 従来手法の全データと提案手法の全データ

この検証には先行研究の手法を用いる必要があるため、小規模データセットを用いて評価を行う。

4.2.2 提案手法の有効性

(2) および (3) で構成される提案手法の有効性を確認する。(2) は以下の 4 種類の分類器を訓練・評価することで検証する。

Dense 密なセグメントのみを用いた 1 つの分類器。

Sparse 疎なセグメントのみを用いた 1 つの分類器。

All (Proposed) 密なセグメントを用いた分類器と疎なセグメントを用いた分類器の組み合わせ。

All (Baseline) 全てのセグメントを用いた 1 つの分類器。

All (Baseline) は従来手法と同様に異なるデータをデータレベルで融合する方法である。他方 All (Proposed) は決定レベルの融合を行う手法である。

(3) は以下の 3 種類の決定を元に評価を行うことで検証する。

No ensemble セグメント単位での判定。

Voting バトル単位での多数決による判定。

Stacking バトル単位でのスタッキングによる判定。

4.2.3 パラメタによる性能変化の確認

パラメタによる性能変化を確認するために、疎密判定の閾値 ρ および セグメント長 θ を変化させた場合の性能を確認する。 ρ は入力密度の閾値であり、大きいほど密なセグメントが少くなり疎なセグメントが多くなる。また θ は大きいほど 1 バトルから得られるセグメントは少なくなる。

それぞれ具体的な値は以下の通りである。

- $\theta = 250$ で固定して $\rho = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ と変化させた場合
- $\rho = 0.5$ で固定して $\theta = 250, 500, 750, 1000$ と変化させた場合

なお、 ρ を変化させる際に $\theta = 250$ としているのは、セグメントを短くし数を増やすことで、 $\rho = 0.2, 0.8$ の場合も十分な学習データを確保するためである。

4.3 評価

本研究ではデータの不均衡に配慮した交差検証手法、評価指標を採用する。本手法ではモデルに one-vs-rest 二値分類器を用いており、1人分のデータが陽性、それ以外のデータが陰性となり量に偏りが生じるためである。

4.3.1 交差検証

いずれの実験においても学習および評価は Stratified Group K-Fold を用いて行う。Stratified K-Fold はデータ全体での正解ラベルの分布を保ちながらデータを分割する手法であり、正解ラベルが偏っている場合に有効である。Group K-Fold はデータがそれぞれあるグループに所属している場合に同一グループのデータが訓練データまたはテストデータのいずれかにのみ含まれるように分割する手法である。本実験ではある試合のデータから複数のセグメントが得られるため、データリーケを防ぐことを目的として同一試合を 1 つのグループとする。

フォールド数を 5 として上記の制約を満たすように分割し、4 フォールドを訓練データ、1 フォールドをテストデータに用いる。訓練データを用いて学習する際には Random Forest のハイパーパラメタを Optuna [9] を用いて最適化した。50 回の試行で平均 F1 スコアを最大化するようパラメタを探査したのち、得られた最適パラメタを用いて訓練データにより再学習することで評価対象のモデルを作成する。

4.3.2 評価指標

評価指標には精度 - 再現率曲線下面積 PR-AUC (Precision Recall Area Under Curve) と等価エラー率 EER (Equal Error Rate) を用いる。PR-AUC は精度と再現率のトレードオフを評価する指標で、不均衡データの性能評価に用いられる。0 % から 100 % の値を取り 100 % に近いほど誤認識の少ないバランスの取れた分類器であることを示す。EER は他人受容率 (FAR : False Acceptance Rate) と本人拒絶率 (FRR : False Rejection Rate) のトレードオフを評価する指標である。低い他人受容率は高いセキュリティを、低い本人拒絶率は高いユーザビリティを示す。そのため、これらが等しくなる点として定義される EER はセキュリティと実用性を両立した認証システムであることを示す、運用上重要な指標である。0 % から 100 % の値を取り 0 % に近いほど高性能を示す。具体的な定義は付録 A.1 に示す。

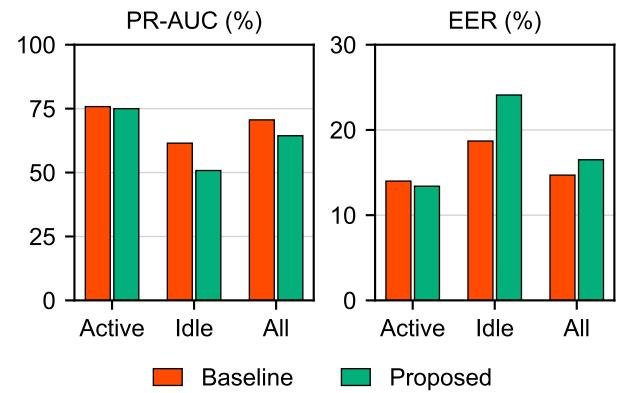


図 2 小規模データセットでの評価。各データの既存手法 (Baseline) と提案手法 (Proposed) での性能を PR-AUC (左), EER (右) で評価した。Active は試合中・密データ、Idle は休憩中・疎データ、All は全データを表し、パラメタは $\rho = 0.5, \theta = 500$ である。

Fig. 2 Evaluation on the small dataset. The identification performance of baseline and proposed methods was measured using PR-AUC (left) and EER (right). Active represents in-game dense data, Idle represents idle sparse data, and All represents all data, with parameters set to $\rho = 0.5, \theta = 500$.

5. 結果・考察

本節では前節で説明した実験の結果を示し、その議論を行う。本研究の目的はコントローラ操作の入力密度に着目したプレイヤー認証手法の提案であり、その有効性を示すことにある。そこでまず入力密度を用いることの有効性を示し、次に提案手法の構成要素、そのパラメタごとの影響について述べる。

5.1 小規模データセットによる疎密分類の有効性

図 2 は小規模データセットを用いて行った、既存手法 [4] と提案手法の比較である。詳細な値は付録 A.2 に示す。休憩中のデータを用いる場合 (Idle および All)，提案手法は既存手法に比べ大きく性能が低下している。一方で試合中のデータのみを用いる場合 (Active)，PR-AUC は 75.8 % から 75.0 % とわずかな低下に留まり、EER は 14.0 % から 13.4 % とわずかに向上している。

この結果は、入力密度に基づく疎密分類の有効性を示している。セグメントの分類に、既存手法は各ラウンドの開始時刻・終了時刻を用いており、提案手法は入力密度を用いてそれを代替していた。これにより抽出された密なデータを用いることで、従来手法とほぼ同等の識別性能を達成できた。この事実はプレイヤー認証においてラウンドの開始・終了時刻のように試合の進行に関する情報が無くとも、コントローラ操作の情報さえあれば認証が可能であることを示唆している。

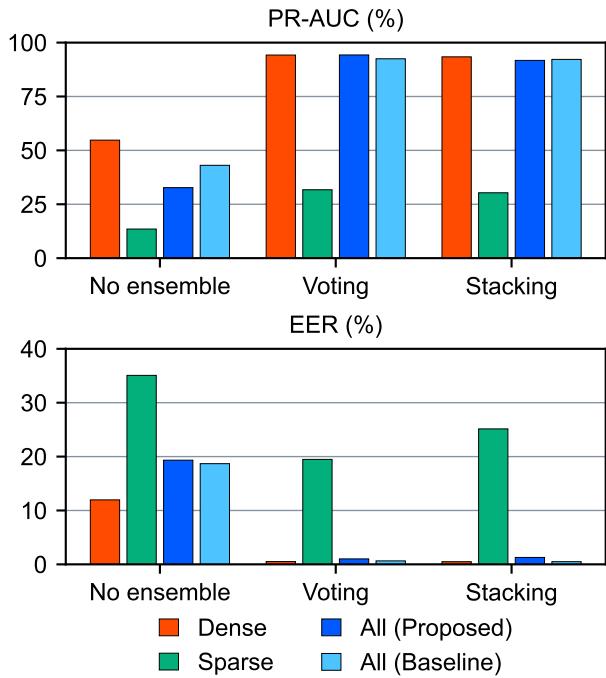


図 3 大規模データセットでの評価. 4 種類のモデル（各バー）と 3 種類のアンサンブル手法（横軸）を組み合わせた識別性能を PR-AUC（上）と EER（下）で評価した. パラメタは $\rho = 0.5, \theta = 500$ である.

Fig. 3 Evaluation on the large dataset. The identification performance was evaluated using PR-AUC (left) and EER (right) for four models and three ensemble methods. The parameters were set to $\rho = 0.5, \theta = 500$.

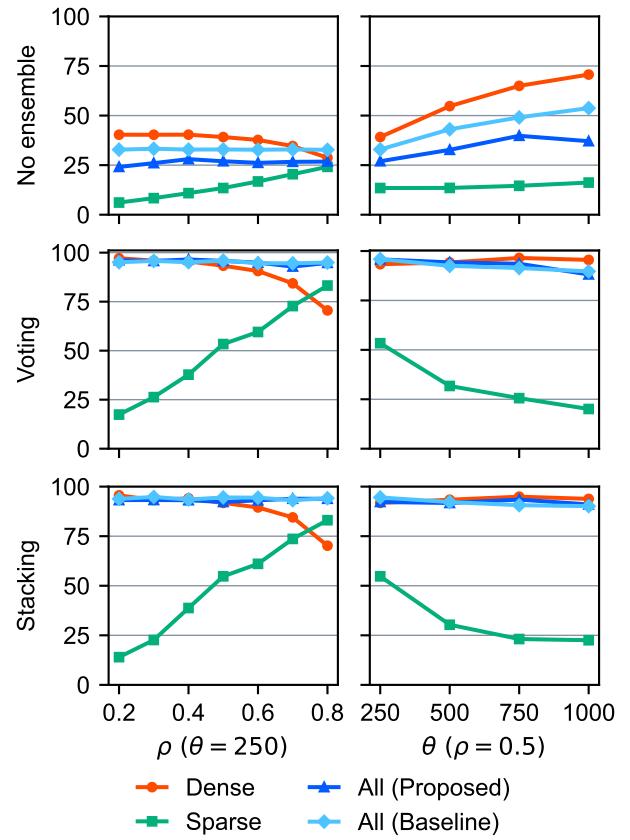


図 4 パラメタ変化に対する性能比較. 各モデル、アンサンブル手法について、パラメタ ρ （左）、 θ （右）を変化させた時の性能を PR-AUC で評価した.

Fig. 4 Parameters and performance. The performance was evaluated by PR-AUC when changing the parameters ρ (left) and θ (right) for each model and ensemble method.

5.2 大規模データセットによる疎密分類の有効性

図 3 に大規模データセットにおける評価の結果を示す. 詳細な値は付録 A.2 に示す. 4 種類のモデルについて、アンサンブル前は Dense が最も性能が高く、All (Baseline), All (Proposed), Sparse と続く. 一方アンサンブル後は Dense, All (Proposed), All (Baseline) が拮抗し、Sparse が最も性能が低い. また 3 種類のアンサンブル方式についても Voting, Stacking により飛躍的に識別性能が向上することがわかる. 最も優れているのは Dense および All (Proposed) を Voting でアンサンブルしたものであった.

モデルごとの結果から、Dense の性能の高さがわかる. Dense はアンサンブル前は最も高い性能を示しているほか、アンサンブル後に 2 つの All の性能が向上しているのは Sparse による誤った分類を補正しているからだと考えられる. この結果は先行研究 [4] の結果とも一致している. 先行研究では試合中のデータが全データより優位なのはデータセットの単調さゆえである可能性が否定できていなかった. しかし本研究の大規模データセットの評価においても、試合中のデータが優位なことが再確認できた.

5.3 パラメタによる性能変化

図 4 は疎密分類における入力密度の閾値 ρ とセグメント長 θ の変化させた時の PR-AUC である. ρ が大きくなると Dense では性能の低下が、Sparse では性能の向上が見られる. また、 θ の増加に対して Sparse では性能の低下、それ以外では性能の向上が見られる.

疎密判定における入力密度の閾値 ρ が大きくなると Dense にはより入力密度の大きいセグメントだけが残る. このことは $\rho < 0.4$ では大きな影響を持たないが、それ以降は個人性を持った入力パターンが失われていき、Dense の性能の低下と Sparse の性能向上をもたらす. このことから、本データセットにおいては入力密度が 0.4 以上のセグメントが個人性を有しているといえる.

また、セグメント長 θ が大きくなると 1 セグメントに含まれるフレームの数は大きくなるが 1 バトルから得られるセグメントは少なくなる. これにより Dense では 1 セグメントあたりの情報が増えることで性能が向上しており、Sparse では情報が増えない一方でセグメント数が減ることで性能が低下している.

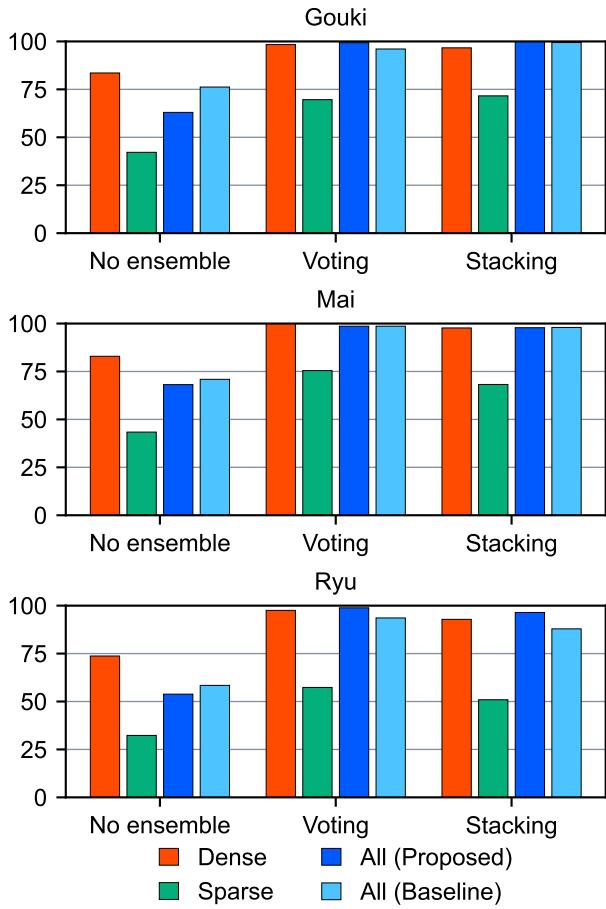


図 5 キャラクター別の評価. Gouki (上), Mai (中), Ryu (下) の 3 キャラクターごとにプレイヤー識別性能を PR-AUC (%) で評価した.

Fig. 5 Evaluation by characters. The player identification performance for three characters, Gouki (top), Mai (middle), and Ryu (bottom), was evaluated using PR-AUC (%).

5.4 キャラクター別の性能

同じキャラクターを使用するプレイヤー同士では入力パターンが類似する可能性がある。格闘ゲームにおいてコントローラ入力は技を構成するボタンの組み合わせに依存し、その組み合わせはキャラクターごとに異なるためである。そこで使用しているプレイヤー数が多い、Gouki (11名), Mai (9名), Ryu (8名) の 3 キャラクターについて、その使用プレイヤー内の識別性能を評価した。

図 5 にキャラクターごとの PR-AUC を示す。アンサンブル前は特に Ryu の性能が低いが、アンサンブル後はいずれのキャラクターも Dense または All (Proposed) で高い PR-AUC を達成している。これより、同一キャラクターを使う場合でもコントローラ操作を用いたプレイヤー認証が可能であるといえる。また図 3 に比べると全体的に性能が高い傾向が見られる。これは全員の 60 名に比べてプレイヤー数が少ないと考えられる。

5.5 本研究の限界

本研究にはいくつか限界がある。まず、本研究では SF6 のコントローラ操作のみを扱ったが、他のゲームにおいても同様の手法が有効であるかは未検証である。入力密度に注目する手法は格闘ゲームのように高頻度で疎密が入れ替わるゲームにおいてのみ有効な可能性がある。その場合、入力密度だけでなくより一般化した判断基準が必要になる。さらに、評価がプレイヤーの識別という基礎的なタスクに焦点を当てており、より複雑かつ実用的な状況への適用は未検証である。これは本研究が試合中データが重要なことをコントローラ操作の情報しかない状況でも確認することを目的としているためである。今後は未知のプレイヤーによるなりすましの検出、時間経過による行動変化への耐性といった、現実に考えられる課題に向き合う必要がある。

6. おわりに

本研究は、e スポーツの公平性担保を背景に、コントローラ操作の記録のみで完結するプレイヤー認証手法を提案した。従来手法は試合の開始・終了時刻という情報を用いていたが、本研究では操作の入力密度に注目した。密な操作が行われるセグメントを抽出し重点的に用いることで、付随情報なしに従来手法と同等の識別性能を達成できることを小規模データセットによる実験で確認した。さらにプレイヤー 60 名、307 試合分の大規模データセットでの実験では PR-AUC 94.22 %, EER 0.52 % という高い識別性能を達成できることを実証した。この結果は、操作が高密度に行われる試合中のデータがプレイヤーの個人性を強く反映するという先行研究の主張を、より大規模で実用に近い環境においても確かめるものである。

今後は 5.5 に示したように他のゲームやデータへの汎化性能の検証と実用的な課題への対応を行う。これらを通してあらゆるジャンルのゲームにおいてプレイヤーの行動に基づく本人確認手法が確立されることで、より公平な e スポーツ環境の実現に寄与できることを期待する。

謝辞 本研究は、JST ムーンショット型研究開発事業 JPMJMS2215 の支援を受けたものです。

参考文献

- [1] Zimmer, F., Irvan, M., Perera, M. N. S., Tamponi, R., Kobayashi, R. and Yamaguchi, R. S.: Player Behavior Analysis for Predicting Player Identity Within Pairs in Esports Tournaments: A Case Study of Counter-Strike Using Binary Random Forest Classifier, *Proceedings of the 58th Hawaii International Conference on System Sciences (HICSS)*, pp. 4280–4289 (2025).
- [2] Irvan, M., Zimmer, F., Kobayashi, R., Perera, M. N. S., Tamponi, R. and Yamaguchi, R. S.: Identifying Esports Players through Unique Sequential Behavioral Signatures in Competitive Fighting Game Matches, Technical Report 124, no. 226 (BioX2024-84), 電子情報通信学会 (2024).

- [3] Siddiqui, N., Dave, R. and Seliya, N.: Continuous User Authentication Using Mouse Dynamics, Machine Learning, and Minecraft, *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pp. 1–6 (2021).
- [4] 川本岳史, Zimmer, F., Perera, M. N. S., Irvan, M., 小林良輔, 山口利恵: 対戦型格闘ゲームにおける待機時間中のコントローラ操作に基づくプレイヤー識別手法の検討, 技術報告 124, no. 440 (BioX2024-98), 電子情報通信学会 (2024).
- [5] Mai, Z., He, Y., Feng, J., Tu, H., Weng, J. and Gao, B.: Behavioral Authentication with Head-Tilt Based Locomotion for Metaverse, *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 949–950 (2023).
- [6] Nguyen, T. T., Nguyen, A. T., Nguyen, T. A. H., Vu, L. T., Nguyen, Q. U. and Hai, L. D.: Unsupervised Anomaly Detection in Online Game, *Proceedings of the 6th International Symposium on Information and Communication Technology*, SoICT ’15, Association for Computing Machinery, pp. 4–10 (2015).
- [7] Qian, X., Sifa, R., Liu, X., Ganguly, S., Yadamsuren, B., Klabjan, D., Drachen, A. and Demediuk, S.: Anomaly Detection in Player Performances in Multiplayer Online Battle Arena Games, *Proceedings of the 2022 Australasian Computer Science Week*, ACSW ’22, Association for Computing Machinery, pp. 23–30 (2022).
- [8] Baltrušaitis, T., Ahuja, C. and Morency, L.-P.: Multimodal Machine Learning: A Survey and Taxonomy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 2, pp. 423–443 (online), DOI: 10.1109/TPAMI.2018.2798607 (2019).
- [9] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M.: Optuna: A Next-Generation Hyperparameter Optimization Framework, *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631 (2019).

付 錄

A.1 評価指標の定義

評価指標 PR-AUC および EER の定義を以下に示す。ある分類器が output した確率 p ($0 \leq p \leq 1$) に対し t ($0 \leq t \leq 1$) を閾値として陽性・陰性を判定する。この時、真陽性 (TP)、偽陽性 (FP)、真陰性 (TN)、偽陰性 (FN) の数をそれぞれ $TP(t)$, $FP(t)$, $TN(t)$, $FN(t)$ とする。この時、Precision, Recall を次のように定義できる：

$$\text{Precision}(t) = \frac{TP(t)}{TP(t) + FP(t)}, \quad \text{Recall}(t) = \frac{TP(t)}{TP(t) + FN(t)}.$$

この時、 t を媒介変数として 0 から 1 まで動かした時に得られる Precision-Recall 曲線と Recall 軸の間の面積を PR-AUC と呼ぶ：

$$\text{PR-AUC} = \int_0^1 \text{Precision}(R) dR.$$

また、他人受容率 FAR、本人拒絶率 FRR は次のように定義される：

$$\text{FAR}(t) = \frac{FP(t)}{FP(t) + TN(t)}, \quad \text{FRR}(t) = \frac{FN(t)}{FN(t) + TP(t)}.$$

表 A.1 小規模データセットでの評価。単位はいずれも %。

Table A.1 Evaluation on the small dataset. All values are in %.

データ種別	PR-AUC (↑)		EER (↓)	
	Baseline	Proposed	Baseline	Proposed
Active	75.8	75.0	14.0	13.4
Idle	61.5	50.8	18.7	24.1
All	70.6	64.4	14.7	16.5

表 A.2 大規模データセットでの評価。単位はいずれも % で、All (Prop), All (Base) はそれぞれ All (Proposed), All (Baseline) を表す。

Table A.2 Evaluation on the large dataset. All values are in %, and All (Prop) and All (Base) refer to All (Proposed) and All (Baseline), respectively.

	Dense	Sparse	All (Prop)	All (Base)
PR-AUC				
No ensemble	54.8	30.3	32.7	43.1
Voting	94.2	31.7	94.2	92.5
Stacking	93.4	13.5	91.7	92.2
EER				
No ensemble	11.9	35.1	18.7	19.3
Voting	0.5	19.5	0.6	1.0
Stacking	0.5	25.1	1.3	0.5

FAR(t) と FRR(t) は t の増減に対してトレードオフの関係にあり、ある $t = t^*$ において FAR と FRR が等しくなる。この時の値を EER と呼ぶ：

$$\text{EER} = \text{FAR}(t^*) = \text{FRR}(t^*).$$

A.2 実験結果の数値

5.1, 5.2 の結果の具体的な数値をそれぞれ表 A.1, 表 A.2 に示す。