

# 卓球映像からの打球の攻守推定

楊 明哲<sup>1,2,a)</sup> 橋本 敦史<sup>2</sup> Jiaxin Ma<sup>2</sup> 本田 秀仁<sup>3</sup> 田中 一敏<sup>2,b)</sup>

## 概要

本研究は、スポーツにおける選手の行動選択を理解するために、卓球における打球時の攻守推定という新しい課題を提案する。攻守推定のために 2 名の卓球経験者による選手の攻守ラベルのアノテーションを実施した。その結果、アノテータ間の正解ラベルの一致率は 0.9 となり、一定の信頼ができることが示唆された。一方で、姿勢情報のみから推定する識別器では、予測精度が 0.63 と低く、挑戦的な課題であることも確認した。

## 1. 序論

意思決定は社会の中で生きる人間にとって重要な営みの 1 つであり、その過程を理解することは重要である。特にスポーツにおいては、複数人が勝利を目指して競合する中で多くの意思決定が行われる。このため、スポーツは意思決定の理解における良い題材となる。

スポーツにおいて、意思決定は行動の選択であり、選択された行動は物理的な結果として表出される。従来のスポーツ分析では、主に選手の行動とその結果を分析に用いている。例えば、球技においては選手が操作したボールの方向や回転の統計的な分析を行うことで戦略的な優位性を確立し勝利につなげる取り組みが多数存在する。

一方で、ボールの方向や回転はあくまでも行動の結果であり、選手自体の意図した結果とは必ずしも一致しない。本研究では、より直接的に選手の意図に近い行動結果の表出を推定することを目指し、熟練者による選手の意図の推定結果を、機械により同様に推定する課題として卓球における打球の攻守推定を提案する。

## 2. データセットの作成

### 2.1 競技とモダリティの選定

対象とする競技は以下の 2 つの理由により卓球のシングルスを選択した。1 つ目に、卓球はネットを挟んだ競技と

いう点である。サッカーやバスケットボールなどの競技とは異なり、卓球は選手同士の接触がないため、それぞれの選手の意図を独立に観測できる。2 つ目に高速かつダイナミックという点である。思考の時間が短いため、意思決定において相手選手の行動による影響が大きく、高い戦略性が要求されるスポーツとなっている。

本研究では、攻守推定を実現するためのモダリティとして映像から得られる姿勢情報を用いる。意図や内面状態を推定する一般的な方法としては、脳波などの生体情報を行う手法がある [6], [8]。しかし、生体情報を扱う際には、センサーやケーブルを取り付ける必要があり、スポーツのような激しい運動を伴う競技では測定が困難である。また、映像全体を利用する場合、一般的に大量の学習データが必要となってしまう。しかし、攻守のように熟練者でないとアノテーションが不可能な推定対象に対して、大規模なデータセットを構築することは難しい。そこで本研究では本課題に対する最初の一步として、映像から選手の意図の表出がある程度反映されると考えられる姿勢情報を入力として攻守判定を行うことを試みる。

### 2.2 正解ラベルの付与

本研究の最終的な目的は、行動選択が勝利という目標に対してどのように影響するかを分析することにある。このため、データセットを作成するにあたり、試合開始から終了まで全てのプレーが撮影されている必要がある。このような映像として市販の全日本卓球選手権の動画を利用した<sup>\*1</sup>。購入した動画のうち、動画内容に偏りがなく、かつ、プレイスタイルが多様になるように男子シングルス 2 本、女子シングルス 2 本、カットマンを含む試合の合計 5 試合を選択して用いる。購入した動画は試合とは関係ないシーンが含まれているため、事前にプレーを行なっているシーンのみを切り出す処理を行う。切り出した後の動画は平均で 390 秒であった。

正解ラベルとしてその行動の意図が「攻めているか (O)、攻めていない (X)」(攻守ラベル)、その行動はどちらの選手のラベルであるか「選手 A (A)、選手 B (B)」(選手ラベル) の組み合わせの 4 つのラベル (OA,OB,XA,XB) を用意

<sup>1</sup> 東京大学

<sup>2</sup> オムロンサイニックス株式会社

<sup>3</sup> 追手門学院大学

<sup>a)</sup> mingzhe-yang@g.ecc.u-tokyo.ac.jp

<sup>b)</sup> kazutoshi.tanaka@sinicx.com

<sup>\*1</sup> [https://www.butterfly.shop/category/ALL\\_JAPAN/](https://www.butterfly.shop/category/ALL_JAPAN/)

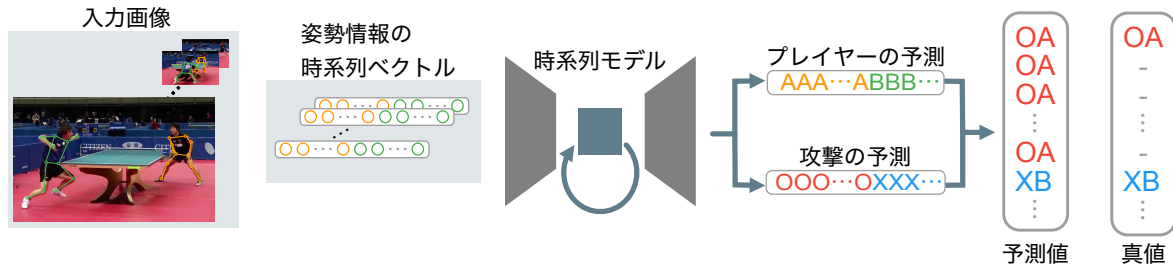


図 1: モデル学習の流れ

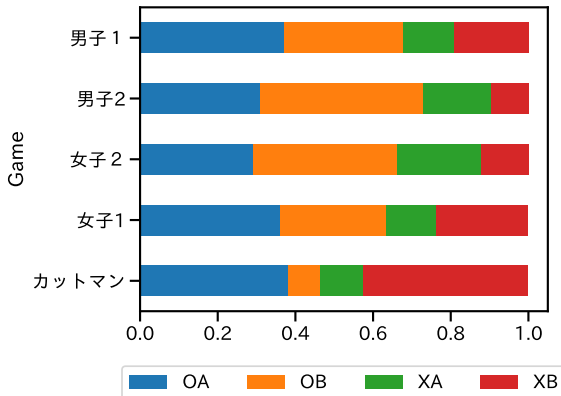


図 2: アノテーションしたラベルの分布。

した。選手 A の決定については、第一ゲームの試合開始時サービス権を有している選手とする。この選手ラベルは各ゲームごとに変化せず、1つの試合を通して一貫している。正解ラベルを付与するタイミングは、選手がラケットにボールがあたる1フレームのみにつける。

今回正解ラベルとして「そのプレーは得点を獲得するために攻めているか（いないか）」を設定している。この攻守ラベルの判断には、卓球のルールだけでなく、戦術に関して深い知識が必要であると考え、アノテータとしては中学生から現在まで選手として卓球に携わっている大学生2名(男女一名ずつ)に依頼した。アノテータには、今回選択した5本の動画とは別の卓球動画に対してアノテーションの練習を行ってもらい、正解ラベルの解釈のすり合わせを行った。

アノテータの付与したラベルの品質を確認するために、アノテータが付与した5試合分のラベルの一致率について確認した。選手ラベルの一致率は、 $0.947 \pm 0.053$ 、攻守ラベルについては  $0.900 \pm 0.045$  であり、4つのラベルのマクロ精度は、 $0.934 \pm 0.04$  であった。選手ラベルの不一致については、アノテータの作業精度による問題だと考える。また攻守ラベルの不一致については、アノテータごとに、不利な状況の時に無理に攻撃している(O)か相手に打たされているか(X)で一致していない状況が複数例確認できた。これはアノテータのこれまでの卓球経験や、男子と女

表 1: 取得した姿勢情報と欠損率とラベルの比率

総フレーム数	ラベルの存在率	両方存在
$9755.2 \pm 6231.2$	$0.068 \pm 0.01$	$0.47 \pm 0.21$

子でプレーの性質が異なることから起因していると考えられる。しかしながら、2人のアノテータのラベル一致度は0.90を超えており、攻守ラベルの判断は少なくとも人間にとっては高い精度で達成可能であることが確認された。

### 2.3 姿勢情報の抽出

卓球の動画から選手の姿勢情報を獲得するために、人物の検出と姿勢予測を同時に行うことができる Kapao ? を利用した。Kapao では同一人物の追跡までは行えないため、物体追跡の手法である DeepSort を組み合わせて用いた [7]。DeepSort では人物に対して ID を割り振ることが可能であるが、選手が画面外に出る時、カメラが切り替わるときなど、追跡結果 (Tracklet) が分割されてしまい、人物に付けられた ID が変化する。このため、手動で Tracklet 間の ID の照合を行い、試合を通じて ID が一貫するように調整した。

これら一連の手続きを行い、獲得した姿勢情報の統計情報について表 1 に示す。Kapao によって映像から姿勢情報を獲得できるが、両方の選手の姿勢情報が取得できている割合は  $0.47 \pm 0.21$  であり、姿勢情報についてはラリー中の半数はどちらかの選手の姿勢情報は欠損している。また全フレームに対して、ラベルが付与されている、つまり、打球の瞬間とされたものは 6.8% となっている。使用した動画は FPS が 25 である、つまりラベルはこれは平均して、0.59 秒にごとに付与されている。

### 3. 識別器の設計

映像から取得した姿勢情報をもとに、攻守推定を行うために、MLP と LSTM ベースの2つのモデルを作成した。識別器の設計の概要を図 1 に示す。

MLP モデルとして、全結合層を3層重ねたモデル構成を採用した。各全結合層の後に、Batch-Normalization, 活性化関数として PReLU を用いる。この出力が次の全結合

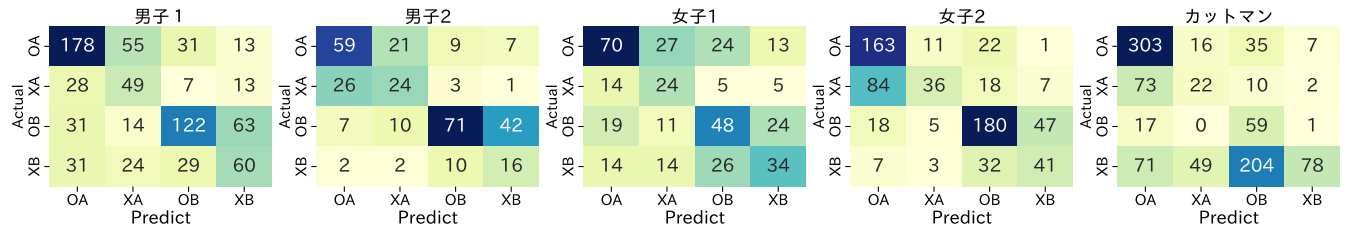


図 3: 5 試合の混同行列のヒートマップ. 右上左下の 4 箇所は選手ラベルの誤分類, 左上右下の 4 箇所は攻守ラベルの誤分類を示す.

層の入力になる.

本研究で入力に用いるデータセットには, 選手の行動について, 時系列が存在している. そこで時系列を考慮するために, LSTM [1], [2] を利用したモデルを作成した. このモデルでは, 入力に対し全結合層を適用し, Batch-Normalize を通し, 活性化関数で PReLU を用いる. その後, LSTM 層を通し, 最後に全結合層によって出力を行う.

過学習の抑制を行うために, LSTM モデル, MLP モデルの両方で, 入力, 出力時の全結合層の直前に Drop out 層(dropout rate = 0.2, 0.5) を追加する.

### 3.1 損失関数

本研究では, モデルの学習を行うために 2 つの損失関数を設定し, マルチタスク問題として扱う. まず 1 つ目のタスクは選手ラベルの二値分類予測である. 2 つ目のタスクは, 攻守ラベルの二値分類予測. この 2 つの二値分類に対して, 損失関数をそれぞれ交差エントロピー誤差を設定し, この二つの損失関数を用いたものが今回最適化すべき損失関数となる.

## 4. 実験

Section2 の手順で作成したデータセットから, 映像から獲得可能な姿勢情報を用いて, 選手の行動意図を推定できるかを検証するために, モデルの学習を行う. データセットの真値としては, ラベルの一致率が 5 試合分のマクロ平均が  $0.934 \pm 0.04$  で高いことから, 片方のアノテータが付与したラベルを真値として学習に用いる.

### 4.1 データの前処理

Kapao で取得した姿勢情報は映像内での絶対位置の情報であるため, 入力データとして扱うためにデータの前処理を行う. まず姿勢情報の正規化を行う. Kapao で獲得した選手の姿勢情報のうち右腰と左腰の midpoint を姿勢情報の原点, 選手の胴の長さを 1 となるように, 各姿勢情報のキーポイントの正規化を行う [3], [5]. データセットの中には図 4 のように, 姿勢情報を獲得できていないフレームが存在している. 姿勢情報が存在しない時は数値を 0 で補完を行った. 卓球の一連のプレーを時系列データとするために,

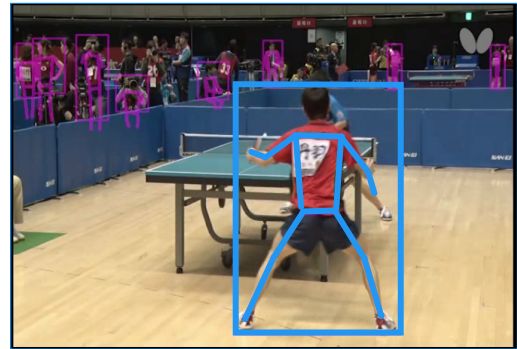


図 4: Kapao で姿勢抽出に失敗している例. 手前の選手に隠れて, 奥の選手の姿勢を取得できていない

固定の時系列長ごとにデータセットを切り出す.

前述した学習に用いるデータセットはサイズが小さく, 過学習をする可能性がある. そこで, 過学習を抑制するために, データ拡張を行う. データ拡張では, ガウシアンノイズの追加による 2 パターンと左右反転の合計 3 パターンを行った. 1 つ目のパターンでは, 選手の体型に対するばらつきをシミュレートする目的で, 系列全体に対して共通のオフセット  $\epsilon \sim N(\mu, \sigma)$  を乱数により生成し, 加える. 2 つ目のパターンでは, 各時刻での関節の検出位置の誤差をシミュレートする目的でオフセット  $\epsilon$  をフレームごとに独立に生成し, 加える. 3 つ目に, 選手の姿勢の向きなど, 撮影映像による制約から起因した過学習の抑制を目的に, 左右変転によるデータ拡張を行う.

### 4.2 学習・評価方法

モデルの学習, 評価を行うために, Leave one-game out を用いる. モデルの訓練時では作成した 5 試合分のデータセットのうち, 4 試合をモデルの訓練に用いて, 残りの 1 試合でモデルの性能の評価を行う. Section3 で設定した損失関数をもとに, 最適化関数として AdamW [4] を用いる. 訓練時の学習率は 0.0001, バッチサイズは 32, 総エポック数は 20 に固定し, これらのハイパーパラメータに関しては, 調整は行わない. 評価指標としては, 選手ラベルと攻守ラベルの組み合わせの 4 クラスのマルチタスク分類の正答率を用いる.

表 2: 20 エポック訓練後の予測精度

model	min	max	mean (std)
LSTM	0.472	0.622	0.536 $\pm$ 0.059
MLP	0.441	0.566	0.515 $\pm$ 0.052

### 4.3 実験結果

表 2 に、Leave one-game out の評価データでの予測精度を示す。5 試合の平均予測精度が MLP モデルでは 0.515 で LSTM モデルでは 0.536 でわずかに LSTM モデルの方が予測精度が高いことが確認できた。5 試合の混同行列のヒートマップを図 3 に示す。ヒートマップを見ると、左上 4 つ右下 4 つの誤分類、つまり攻守予測の分類において誤分類が、選手の誤分類と比べて多い傾向が確認できた。特にカットマンのいる試合を見ると、攻守ラベルの予測において誤分類が多いことが確認できた。これは訓練データにはカットマンのような守備よりの戦型が含まれておらず、訓練データと検証データのラベル分布が異なっているからと考えられる。

本研究では、予測モデルの学習で、選手ラベルの予測と攻守ラベル予測の 2 つを用いて、学習を行ってきた。この 2 つの損失関数が結果にどのような影響を与えるかを見るために、使用していた 2 つの損失関数を個別に取り除いた場合の結果について調査する。まず選手ラベルの予測のみについて実施する。攻守ラベル予測の損失関数を削除することで、映像から取れる情報から、選手ラベルの予測を行う。実験結果は表 3 に示す。選手ラベル予測では、精度の平均が 0.809 であった。アノテータの選手ラベルの一致度が 2 より、0.947 であることと比較すると、学習モデルがどちらの選手か予測するタスクは比較的容易であることが分かる。誤分類した例を見ると、ネット付近のプレーで選手の動きが小さいものが確認できた。動きが小さいため、姿勢情報だけでは、選手が打球したかどうかか予測困難であったことが考えられる。

次に攻守ラベル予測のみについて実施する。選手予測の損失関数を削除することで、映像からそのプレーが攻めているかの予測が行われる。実験結果は表 3 に示す。攻守ラベルの予測に関しては精度の平均が 0.631 であり、選手ラベル予測よりも精度が低く難しいことが確認できた。アノテータのラベルの一致度が、 $0.900 \pm 0.045$  であったことを考慮すると、映像から得られる情報をもとに、機械学習モデルは人間が見て判断するよりも難易度が高いことが分かった。

## 5. 結論

本研究では、卓球の行動意図の推定を試みた。経験者の視点から「その行動は得点のために攻めているか」の新たなラベルを設定し、打球の攻守推定という新しい課題を提案した。経験者によるラベル一致度は高かったが、現在の

表 3: 片方のタスクに限定した時に学習した時の予測精度。モデルは LSTM モデルを用いている。

タスク	min	max	mean (std)
選手・攻守予測	0.472	0.622	0.536 $\pm$ 0.059
選手予測	0.780	0.853	0.809 $\pm$ 0.037
攻守予測	0.562	0.702	0.631 $\pm$ 0.065

手法やデータセットを用いた機械による推定精度はこれに匹敵するまでには至らなかった。人間が行動から意図を理解する能力は、現状の技術では機械には難しい可能性が示唆された。ただし、この困難さは現状の手法やデータセットのサイズ、姿勢以外の情報の必要性に起因する可能性がある。今後、このデータセットや課題設定を中心として様々な取り組みを行っていききたい。

## 謝辞

本研究は JSPS 科研費 JP21K18568, および、JST さきがけ JPMJPR22C6 の助成を受けたものです。

## 参考文献

- [1] Gers, F. A., Schmidhuber, J. and Cummins, F.: Learning to forget: continual prediction with LSTM, *Neural computation*, Vol. 12, No. 10, p. 2451–2471 (2000).
- [2] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Comput.*, Vol. 9, No. 8, pp. 1735–1780 (1997).
- [3] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft COCO: Common Objects in Context, *Computer Vision – ECCV 2014*, Springer International Publishing, pp. 740–755 (2014).
- [4] Loshchilov, I. and Hutter, F.: Decoupled Weight Decay Regularization, *International Conference on Learning Representations* (2019).
- [5] Newell, A., Yang, K. and Deng, J.: Stacked Hourglass Networks for Human Pose Estimation, *Computer Vision – ECCV 2016*, Springer International Publishing, pp. 483–499 (2016).
- [6] Patel, A. N., Howard, M. D., Roach, S. M., Jones, A. P., Bryant, N. B., Robinson, C. S., Clark, V. P. and Pilly, P. K.: Mental state assessment and validation using personalized physiological biometrics, *Frontiers in human neuroscience*, Vol. 12, p. 221 (2018).
- [7] Wojke, N. and Bewley, A.: Deep Cosine Metric Learning for Person Re-identification, *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 748–756 (2018).
- [8] 松居辰則: 生体情報を用いた学習者の心的状態推定と学習支援の試み, *教育システム情報学会誌*, Vol. 36, No. 2, pp. 76–83 (2019).