

卓球の得点予測における重要要素の分析

楊 明哲^{†,††} 橋本 敦史^{††} Jiaxin Ma^{††} 本田 秀仁^{†††} 田中 一敏^{††}

[†] 東京大学

^{††} オムロンサイニックス

^{†††} 追手門学院大学

E-mail: [†]mingzhe-yang@g.ecc.u-tokyo.ac.jp

あらまし スポーツのデータ分析において、機械学習モデルを用いた勝敗予測や戦術解析が盛んに行われている。サッカーなどを題材とする既存研究では、試合全体を考慮した戦略に注目が集まっている。一方で、卓球のような一定の得点獲得を目指す競技では1つ1つのプレーに焦点を当てた分析が求められている。そこで本研究では、試合データをもとに各得点における重要な要素を推定することを試みた。具体的には、外部から観測可能な得点状況などの試合データをもとに機械学習モデルによる重要特徴量の推定を行い分析する。さらに本研究では、選手の行動意図を反映した攻守ラベルを卓球経験者によって作成し、得点獲得にどのように寄与するか分析を行った。

キーワード

1 はじめに

スポーツのデータ分析において、機械学習を用いることが盛んに行われている。テニスやサッカーなどを対象とした既存研究では、試合における勝敗予測や重要なイベントの推定が行われてきた[2], [3]。これらは外部から観測可能な情報であるゲームや試合における得点やサーブ権を特徴量として、勝敗予測を行なった。これらの研究では、勝敗予測の重要な要素として、得点数の多さやレシーバーによる得点率の高かさなどを挙げた[3]。これらのスポーツにおける重要要素の推定は、いずれも試合全体に対する戦術アドバイスであった。卓球のように一定の得点の獲得を目指す競技では、一打一打のプレーに対する戦術的アドバイスが重要である。一打一打に注目した分析を行うために、卓球競技における既存研究では、飛行特性、回転特性や配球特性など物理的な球に対する試みは行われている[4]。卓球競技において外部から観察できる情報に基づいた分析は行われているが、外部から直接観測ができない情報を用いた分析は行われていない。そこで本研究では、各プレーに注目した特徴量として、選手の内部状態に注目した攻守ラベル[5]を用いる。攻守ラベルとは、選手がそのプレーを行う際に、得点するために攻撃的なプレーをしたかどうかである。本研究では、卓球経験者による選手の行動に対して、付与した攻守ラベルを用いることで、各プレーにおける重要要素の推定を行った。

2 データセット

2.1 競技の選定と試合状況のデータセット

本研究では、対象の競技として卓球を採用した。卓球は、一定の得点獲得を目指すため一打一打のプレーに対して、戦術的アドバイスが重要である。また、卓球は、サーブによってプレーの流れが変化する競技である。このような特徴を持つ

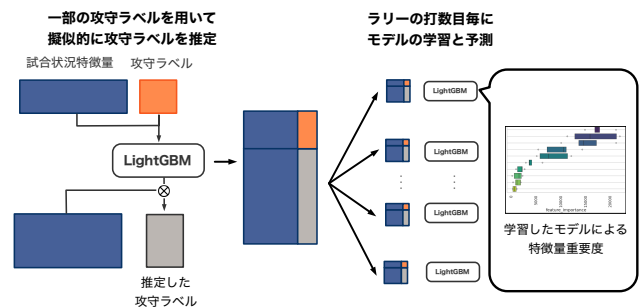


図 1: 本実験における全体の流れ。初めに一部の攻守ラベルを教師として、分類器の学習を行う。その後学習された分類器によって疑似的に攻守ラベルを作成する。次にデータセットを打球目について分割し分類器の学習を行う。

卓球を対象とすることで、各プレーに対する重要要素の推定を行うことができる。本研究では、平成 27 年度および 28 年度の全日本卓球選手権ジュニア男子、全日本選手権男子、女子のシングルス、合計 45 試合に対してデータセット作成を行った。データセット作成には、上記の試合映像をもとに、試合におけるサーブ権、現在の得点状況などの特徴量を人間の手作業によって収集した。表 1 に本研究で用いた特徴量についてまとめる。各特徴量は、すべてサービスを行う選手を起点としたものに特徴量変換を行っている。これによって、分析する際にサーバー、レシーバーの視点から重要要素の分析が可能になる。

2.2 攻守ラベルの付与

本研究では、得点に向けたプレーの重要要素の分析を行うために、試合状況の特徴量に追加して、選手のプレーに対する攻守ラベルの利用を考える。本研究では攻守ラベルを、選手のプレーの意図が得点のために攻めているか否かと定義した。この攻守ラベルの判断は、卓球のルールだけでなく、戦術に関する深い知識が必要だと考え、ラベル作成には卓球経験者を用いた。

表 1: 作成した特徴量一覧。サーバーを S、レシーバーを R とする。

特徴量	内容	変数タイプ
Set	試合中の何ゲーム目か	数値
Play	同ゲーム内で何本目のラリーか	数値
NRallyOrdered	現在の打球がラリー中の何打目か	数値
ScoreStatusFromServer	S の現在の得点	数値
ScoreStatusFromReciever	R の現在の得点	数値
Server	現在の打球が S のものか否かの 2 種類	カテゴリカル
ScoreStatusDiffBinaryFromServer	S と R の得点差があるかどうか。1) S のほうが得点が高い 2) R のほうが得点が高い 3) 同点の 3 種類	カテゴリカル
AttackLabelsFromServer	S に攻撃意図があるか否か、R に攻撃意図があるか否かの 4 種類	カテゴリカル

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 1 名ずつ) に依頼した。アノテータには、利用した試合データのうち、5 本の動画に対してアノテーション付与を行った。アノテーション対象動画を 5 本のみにした理由としては、今回の作成したラベルは選手の一打ごとに攻守ラベルを付与するものであり、すべての動画に対してラベル作成するのは非常にコストがかかるためである。アノテータが付与したラベルの品質を確認するために、アノテータ同士のラベル一致率を計算した。アノテータによるラベル一致率は、 0.934 ± 0.04 であった。アノテータのラベルの不一致は、アノテータの作業精度や、アノテータのこれまで経験に依存しているために発生している問題であると考ええる。しかしながら、ラベルの一致度が 0.9 を超えていることから、少なくとも人間による攻守ラベルは高い精度で達成可能であると考ええる。

本研究では、アノテーションコストの観点から、データセットで用いる一部の試合のみにしか攻守ラベルが存在していない。そこですでに付与されている攻守ラベルを用いて、擬似的に攻守特徴量の作成を行なった。このアプローチは半教師あり学習に類似している。攻守ラベルが付与されている 5 試合のデータセットを訓練データとして、LightGBM [1] による攻守ラベルの擬似特徴量の生成を行った。LightGBM は、勾配ブースティング木を用いた機械学習モデルであり、高い精度で予測を行うことができる。

3 実 験

3.1 分類器の学習と評価方法

本実験における全体の流れを図 1 に示す。Section2 で作成した特徴量を用いて、各プレーにおける得点を予測し、得点獲得に重要な要素の分析を行う。さらに本研究にて新たに作成した攻守ラベルはプレーの重要度にどのような影響を与え、戦術分析に有用であるかを検証する。分類器として LightGBM を用いた。LightGBM を用いた理由は、LightGBM は、勾配ブースティング木を用いた機械学習モデルであり、特徴量の重要度を計算することができるためである。この特徴量重要度は、各特徴量が予測に対してどの程度寄与しているかを表す。

作成したデータセットを試合単位で分割し、試合単位で交差

表 2: 攻守ラベルありの場合となしの場合の精度比較

訓練データセット	攻守ラベルなし	攻守ラベルあり
1 打球目 (サーブ)	0.660	0.731
2 打球目 (レシーブ)	0.621	0.733
3 打球目 (三球目攻撃)	0.625	0.675
4 打球目	0.635	0.695
5 打球目以降の奇数打目	0.489	0.633
6 打球目以降の偶数打目	0.670	0.637

検証を行った。試合単位で交差検証を行うことで、試合単位での汎化性能を評価することができる。また試合単位での交差検証を行うことで、試合単位での特徴量の重要度を計算することができる。本研究では、選手の一打一打に注目して各プレーにおける重要要素の分析を行うため、我々はサーブから打数に応じたデータセットを分けて、それぞれ異なるモデルの学習を行った。本実験では 1 ラリー内の打数に応じて、6 種類に分割した。分割したデータセットはそれぞれ 1) 1 打目 (サーブ)、2) 2 打目 (レシーバ)、3) 3 打目 (三球目攻撃)、4) 4 打目、5) 5 打以上の奇数打 (サーバーの打球)、6) 6 打以上の偶数打 (レシーバーの打球) とした。本研究の正解ラベルは、1 つのラリーにおいて、サーバーとレシーバーのどちらが得点するかを二値ラベルを設定した。サーバーが得点した時は 0 で、レシーバーが得点した時は 1 となる。分類器には試合状況のデータセットが与えられて、そのラリーにおいてサーバーが得点するかレシーバーが得点するかを予測を行う。モデル性能を評価する指標として二値ラベルの正解率を用いた。

4 結 果

Section2.2 で推定した攻守ラベルの正解率について報告する。事前に人間の手で作成した攻守ラベルを正解ラベルとし、交差検証で汎化性能を評価したところ、予測精度は 0.31 であった。

表 2 に、攻守ラベルを用いた場合と用いない場合の各打球における予測精度を示す。攻守ラベルを用いた場合は用いない場合と比べて予測精度が向上していることがわかる。このことから、機械学習モデルで得点を予測する際に攻守ラベルを用いることは精度向上に寄与することが明らかになった。

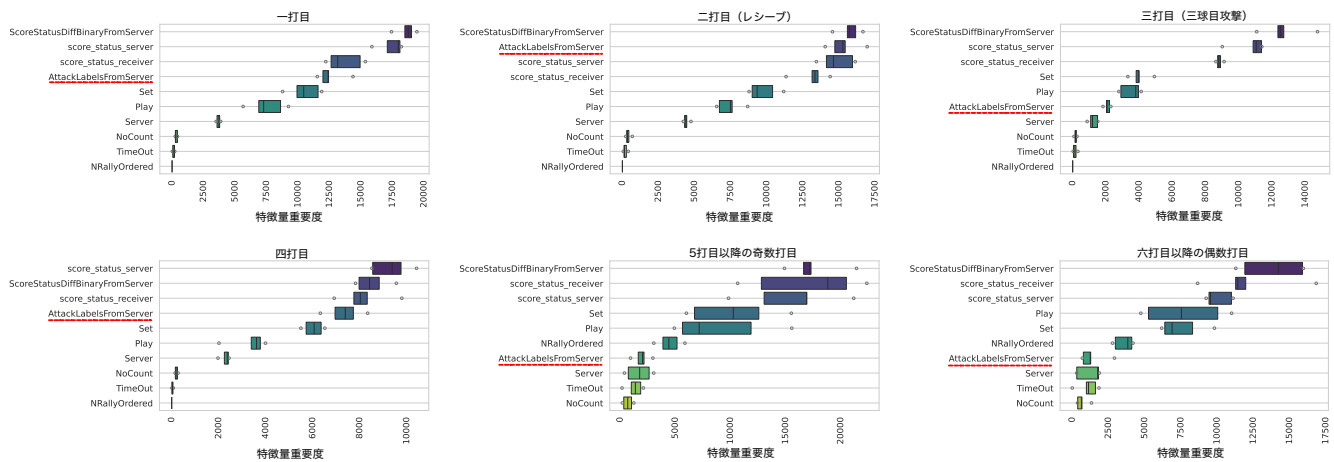


図 2: 分割したデータセット毎に学習した分類器の特徴量重要度。赤点線は、本研究で作成した攻守ラベルを指している。

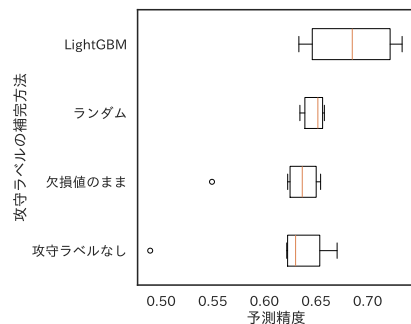


図 3: 攻守ラベルの作成方法の違いによる分類精度の比較

図 2 に、LightGBM での特徴量重要度を示す。分類器として重要な特徴量として、サーバー、レシーバーの得点状況 (ScoreStatusFromServer, ScoreStatusFromResiver)、そして両者の得点差 (ScoreStatusDiffBinaryFromServer) が挙げられていた。これらの特徴量は既存研究でも重要な要素として考えられており、本研究でも同様の結果が示された [3]。攻守ラベルについては、特に 2 打目や 4 打目において、分類器の予測において重要要素になっていることが確認できた。この 2 打目、4 打目はレシーバーによる打球であり、攻守ラベルが重要要素になっていることから、レシーバーによる得点獲得において、この得点獲得のために攻撃する意図が重要であると考えられる。卓球競技をはじめとするネット型競技の性質上、サーブは他者の影響を受けないためサーバーは得点する上で非常に有利である。また卓球は 3 球目攻撃と呼ばれるほど、サーバーはサーブから 3 球目攻撃の展開を作ることが有用な戦術である。レシーバーはこの展開を阻止するためにレシーブを行う。攻守ラベルが重要要素になっていることから、不利な状況になりやすいレシーバーは、得点獲得のために攻撃的なプレーを行うことが重要であると考えられる。

図 3 に、擬似攻守ラベルの生成方法の違いによる予測精度の変化を示す。比較対象として、攻守ラベルを欠損値として扱う方法とランダムに割り当てる方法を用いた。欠損値とランダム割り当てのいずれも分類器を用いて推定した本手法に精度が及ばなかった。本実験において分類器によって付与された攻守ラ

ベルは、予測精度が 0.31 であり、チャンスレートの 0.25 と比べて僅かに高いものであった。ランダムや欠損値と比較すると、分類器による割り当てを行うことで、予測精度の向上が見られた。このことから、さらに精度の高い攻守ラベルを分類器に与えることで、より高い予測精度が期待できる。

5 おわりに

本研究では、卓球のプレーに対する得点予測を行うときに、選手の攻撃意図を表す攻守ラベルを用いることで、分類器の予測精度向上に寄与することを示した。また、分類器の特徴量重要度を見ることで、不利な状況になりやすいレシーバーが得点するためには、選手の攻撃意図が重要であることを示唆した。今後は、既存の選手の配球戦略と攻守ラベルを組み合わせることで、より有用な戦術分析を行うことができると考える。

謝 辞

本研究は、JST、さががけ、JPMJPR22C6 の支援を受けたものである。

文 献

- [1] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 2017 International Conference on Neural Information Processing Systems*, 2017.
- [2] Michal Sipko and William Knottenbelt. Machine learning for the prediction of professional tennis matches. *MEng computing-final year project, Imperial College London*, Vol. 2, , 2015.
- [3] Chen Yang, Tian Yubo, and Zhong Yi. Final project report: Real time tennis match prediction using machine learning. 2017.
- [4] 塩入彬允, 牛山幸彦, 佐藤悠樹. 卓球競技における得点に影響する配球の分析方法の検討. *スポーツパフォーマンス研究*, Vol. 14, pp. 234–242, 2022.
- [5] 楊明哲, 橋本敦史, Ma Jiaxin, 本田秀仁, 田中一敏. 卓球映像からの打球の攻守推定. In *Proceedings of the 2023 Meeting on Image Recognition and Understanding*, 2023.