

ラリー情報を用いたテニスにおけるポイント取得のための重要要素取得

牧野 光秀†

小高 知宏†

黒岩 丈介†

諏訪 いずみ†

白井 治彦‡

† 福井大学工学研究科

‡ 福井大学工学部

1 はじめに

テニスにおいて機械学習を用いた勝敗予測, 戦術解析が行われており, 先行研究では試合の勝利予測, 勝利のための重要な要素の推定が行われた [1].

試合全体のセット間, 試合全体からみられる情報を特徴量とし勝敗予測を行った研究 [2] は行われ, 得られた重要な要素は「ポイントの取得総数が多い」, 「ブレークポイントの取得率が高い」等であった. これらは勝利のために重要な要素ではあるが戦術的なアドバイスには繋がらないという問題点がある. これは用いられた特徴量がポイント間の過程に着目されていないためである.

また速度, 選手位置等のショット情報からポイント取得予測を行う研究 [3] も行われている. しかしラリー間のショット同士の組み合わせ, 割合等が考慮されておらずその点から戦術的な情報を取得できると考えた.

そこで本研究ではラリー間の選手間のショット情報から得られる各ポイントの過程を表す特徴量を用いることでポイント取得のための戦術的に有効な要素を推定できると考えた. テニスシングルス男子プロのポイント間のラリー情報から特徴量を作成し, 戦術的に有効な特徴を推定するのが本研究の目的である.

表 1: 球種の詳細

球種	記号	球種	記号
フォアハンドトップスピン	f	バックハンドトップスピン	b
フォアスライス	r	バックハンドスライス	s
フォアボレー	v	バックボレー	z
フォアハンドスマッシュ	o	バックハンドスマッシュ	p
フォアハンドドロップ	u	バックハンドドロップ	y
フォアハンドドライブボレー	j	バックハンドドライブボレー	k
フォアハンドロブ	l	バックハンドロブ	m
フォアハンドハーフボレー	h	バックハンドハーフボレー	i
左方向へのショット	1	ワイドへのサーブ	4
センター方向へのショット	2	ミドルへのサーブ	5
右方向へのショット	3	センターへのサーブ	6
コート前方へのリターン	7	コート中央へのリターン	8
コート後方へのリターン	9		

表 2: ショット情報の詳細

	サーブ	リターン	他のショット
球種	1st,2nd	f,b,r,s	全て
コース (左右)	4,5,6	1,2,3	1,2,3
コース (前後)	なし	7,8,9	なし
球速	なし	なし	なし
プレーヤーの位置情報	なし	なし	なし

2 方法

本研究の方法としてデータセットからポイント間の過程を示す特徴量を作成し, 解釈性の高い機械学習モデルで学習を行いラリー情報からそのポイントの勝者予測を行う. 学習された機械学習モデルから勝利に大きく影響を与える特徴量を抽出し考察する.

2.1 データセット

データセットには MatchChartingProject[4] を用いる. MatchChartingProject は 1970 年から 2018 年の内 2282 試合, 約 320000 ポイントのラリー情報が記載されている. Hawkeye のように機械で取得されたデータではなく人間の手作業によって作成されている. そのため Hawkeye データと比較して選手位置, ショットの着弾点, 打点の座標, 速度は無く各ショットに関して球種, コースなどが記載されている. 以下表 1 に球種・コース, 表 2 に扱うショット情報の詳細を記載する.

2.2 特徴量作成

上記のデータを用いてラリー開始からラリー終了直前のポイントまでのラリー間のショットデータを用いてポイント間の過程を示す特徴量を作成する. 作成した特徴量を以下表 3 に記載する.

切り返しとは直前の相手のコースに対して自分が逆のコースにショットをすること, 振り回しとは自分の一つ前のショットに対して逆のコースにショットすることとする. 相手を動かす事が有利とされるテニスにおいて有効であると考え.

2.3 学習方法

得られた特徴量に対して機械学習及び特徴量選択を行いポイント取得の重要要素取得を行う.

Acquisition of important elements for point acquisition in tennis using rally information

†Mitsuhide Makino †Tomohiro Odaka †Jousuke Kuroiwa

†Izumi Suwa ‡Haruhiko Shirai

†Graduate School of Engineering, University of Fukui

‡Faculty of Engineering, University of Fukui

表 3: 作成した特徴量の詳細

	特徴量数	内容
サーブ種類+コース	12	カテゴリ
リターン種類+コース	36	カテゴリ
サーブに対するリターン球種+コース	432	カテゴリ
各ショット(球種のみ)の確率, 回数	64	数値
各ショット(球種+コース)の確率, 回数	192	数値
ボレー, ドロップショットをしたか	4	カテゴリ
ショット A(球種+コース) に対する 返球ショット B(球種+コース) の確率, 回数	576	数値
切り返しの確率, 回数	140	数値
振り回しの確率, 回数	140	数値

使用する機械学習モデルはランダムフォレストを用いる。ランダムフォレストはアンサンブル学習アルゴリズムの一つでランダムに選択された説明変数を用いて構築した決定木を弱学習機とし、それらの多数決によって判別を行う。ランダムフォレストの機能として各説明変数の判別に対する影響度を推定することができる。本研究ではそれを利用しポイントを取るための重要な要素を取得する。

作成したラリー開始からラリー終了の1つ前までのショットデータから得られた特徴量を説明変数、特定の選手がポイント取得したかどうかを目的変数として学習を行い、ポイント取得予測、学習機からポイント取得のための重要特徴量の取得を行う。その際条件を変化させる事で予測率、重要特徴量の変化を観察する。

3 実験及び結果

作成した特徴量に対して条件を変え実験を行った。実装には Python3 を使用し、学習には機械学習ライブラリ Scikit-learn を用いた。不順度はジニ係数を用い、10 分割交差検証を行うことで決定木の深さ、個数に対してパラメータチューニングを行った。

対戦条件を変化させサーバーのポイント取得の予測、重要特徴量抽出を行った。表 4 に実験条件及び予測結果、図 3 にフェデラー対ジョコビッチの際に得られた上位 3 位までの重要特徴量と重要度を記載する。

表 4: 実験条件及び予測率

試合条件	S フェデラー R ジョコビッチ	S ナダル R ジョコビッチ	S フェデラー R ガスケ
総ポイント数	3643	4297	564
S ポイント数	2550	2700	394
R ポイント数	1093	1158	170
正解率	0.652	0.640	0.735
F 値	0.755	0.753	0.84

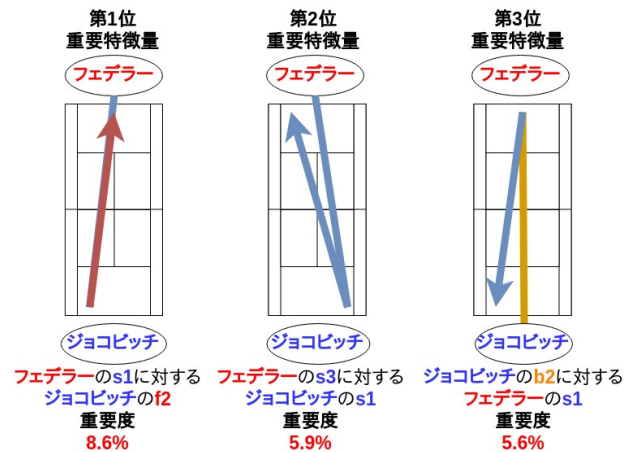


図 1: フェデラー対ジョコビッチ間の重要度及び特徴量

4 考察・まとめ

対戦する選手毎に重要特徴量の差異があることが確認できた。フェデラー対ガスケの試合が他の試合と比較して平均ポイント率、正解率ともに大きく差が出たのは今回実験を行った他の選手と比較してガスケがランク下位の選手でありポイント取得率の少なさに起因すると思われる。今回ランダムフォレストで特徴抽出を行ったが S 側 R 側どちらのポイント寄与に影響を与えているかわからない点があり他のアルゴリズムを用いることでより詳細に影響を確認できると考える。今回はコートサーフェイスを考慮しなかったが選手間で得意なサーフェイス等があり勝率、ショットの選択に大きく寄与すると思われる今後検討する。

参考文献

- [1] Sipko, Michal, and William Knottenbelt. "Machine learning for the prediction of professional tennis matches." MEng computing-final year project, Imperial College London (2015).
- [2] Learning, Machine. "Final Project Report: Real Time Tennis Match Prediction Using Machine Learning." (2017).
- [3] Wei, Xinyu, et al. "The thin edge of the wedge: Accurately predicting shot outcomes in tennis using style and context priors." Proceedings of the 10th Annu MIT Sloan Sport Anal Conf, Boston, MA, USA. 2016.
- [4] https://github.com/JeffSackmann/tennis_MatchChartingProject