多様化するスポーツデータの特徴抽出と可視化

Feature Extraction and Visualization of Diversifying Sports Data

5117E029-2　和田　滉平　　指導教員　河合　隆史　教授

WADA Kohei　　Prof. KAWAI Takashi

概要：　スポーツのデータ分析において，統計学や機械学習の手法を取り入れることによる複雑な選手の能力評価やチームの戦術評価が期待されている．本研究ではフェンシング・バスケットボール・サッカーの 3 種類のスポーツの試合から得られた実測データを対象とし，統計学や機械学習の観点からデータ分析を行うことによって選手やチームの特徴を理解しやすい形で可視化することを目的とする．フェンシングの分析では映像データを対象とし，突きと前後移動に着目して選手の特徴を表す指標を示した．バスケットボールの分析では行動履歴データを対象とし，タイムアウト前後のプレーに着目して各チームが捉える試合の流れを可視化する手法を提案した．サッカーの分析ではトラッキングデータを対象とし，選手の位置関係や速度に着目してパスの成功確率を予測するモデルを作成した．また作成モデルを応用することで，スペースや選手の評価が可能であることを示した．

キーワード：スポーツ統計，機械学習，フェンシング，バスケットボール，サッカー

Keyword: Statistics in Sports, Machine Learning, Fencing, Basketball, Soccer

1　はじめに

スポーツのデータ分析において統計学や機械学習の手法を取り入れることで，それまで力学や幾何学の観点だけでは困難であった，複雑な選手の能力評価やチームの戦術評価などが期待されている[1]．またそれらの複雑な分析結果を受け取る対象の多くはデータ分析に詳しくない選手や観客などであると考えられ，抽出された特徴の表現方法も重要となる．そこで本研究ではフェンシング・バスケットボール・サッカーの 3 種類のスポーツの試合から得られた実測データを対象とし，統計学や機械学習の観点からデータ分析を行うことによって選手やチームの特徴を理解しやすい形で可視化することを目的とする．

2　フェンシングの分析

フェンシングにおいては突きのタイミングと間合いの取り方に着目し，強い選手の特徴を明らかにすることを目的として分析を行う．使用したデータは 2018 年ワールドグランプリ大会 3 大会におけるエペの試合を撮影した映像データ 3 試合分であり，それらに共通して出場する当時の世界ランキング 7 位の NOVOSJOLOV 選手に着目して分析を行った．

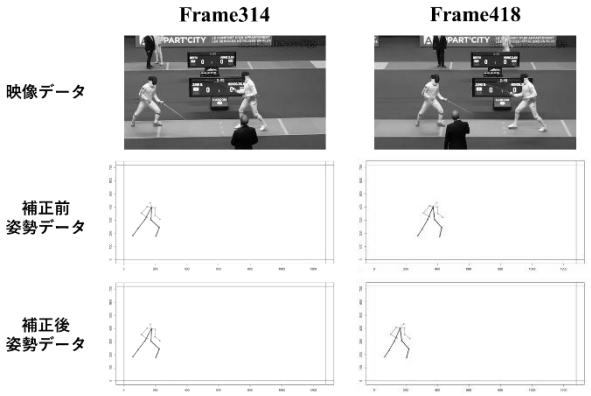

図 1 補正後データの例

2.1　映像を基にした姿勢データの取得

まず試合を得点毎に分割し OpenPose を用いてフレーム毎に画面内に写る人体の 2 次元姿勢座標を推定した後，平均二乗誤差を用いて同一選手のデータのフレーム間の対応付けを行った．またテンプレートマッチングを用いて画面内に映る固定物の座標移動を取得し，これを引くことでカメラ移動による影響の補正を行った(図 1)．

2.2　戦況の定義と選手の特徴

続いて取得した姿勢データを基に戦況を表す指標を定義する．まず肩の角度と肘の角度がともに 135 度以上であり，肩の角速度が正の値である区間を突きの区間とし，また相手の後退(前進)に対し，より大きく前進(後退)できている状況を「優勢な前進(後退)」とする．これを基に突きのタイミングに対して優勢な前後移動を行っているフレームの割合をそれぞれ算出し，特に 1)自身の突き前・突き中に自身の前進が優勢な割合 2)自身の突き後に自身の前進が優勢な割合 3)相手の突き前に自身の後退が優勢な割合 4)相手の突き中・突き後に自身の後退が優勢な割合，の 4 点を戦況を表す指標として定義する．この指標を 3 試合それぞれに対して対戦選手間で比較したところ，本研究で強い選手の一例として着目した NOVOSJOLOV 選手の特徴は，相手の突きの最中及び突きの前後 10 フレームで優勢な後退が出来ていることであると示された．

3　バスケットボールの分析

バスケットボールにおいては各チームが捉える試合の流れを実際のプレー内容を用いて可視化することを目的として分析を行った．分析に使用したデータは 2016-2017 シーズン B1 リーグ戦，全 18 チーム全 540 試合分の行動履歴データである．

3.1　タイムアウトの特徴

まず試合の流れに関わる要素として，悪い流れを変えたい場合や新しい流れを作りたい場合などに取得されることが多いと考えられるタイムアウトに着目し基礎的な分析を行った．この結果，タイムアウト取得時の時間帯や取得時の状況，取得率にはチーム毎に異なる特徴があることが示された．

3.2　タイムアウト前後のプレーにおけるチームの特徴

そこでタイムアウトを取得する要因になると考えられるタイムアウト前のプレーから，タイムアウト中の指示や休息によって変化する可能性が高いと考えられるタイムアウト後のプレーまでを対象として系列パターンマイニングを適用することで，具体的なプレーを用いた試合の流れの可視化を試みた．また系列パターンマイニングの問題点であるイベント間の時間幅を考慮できない，という点に対しては時間情報を付加したデータを作成することによって対処した．系列パターンマイニングによってチーム毎に発生頻度の高いプレーのルールを抽出し特徴を考察した結果，例えば千葉ジェッツは 3P シュートに失敗し相手にリバウンドを取られることを悪い流れとして捉えている可能性がある一方で，富山グラウジーズは自チームのファウルから相手にフリースローを決められることを悪い流れと捉えている可能性があるなど，各チームで異なる特徴を示すことが明らかとなった．

4　サッカーの分析

サッカーにおいては選手の位置関係や移動速度に着目し，パスとスペースに関する分析を行った．使用したデータは 2017 年 J1 リーグ戦の最終 5 節全 45 試合を対象としたトラッキングデータ及び行動履歴データである．

4.1　パス成功確率予測モデルの学習

まず各選手がパスの成功/失敗に与える影響力を表す指標を作成するため，選手の位置や速度などに着目して 6 種の条件を設定する．続いてそれぞれの条件において実際にパスが取得された位置を計算し，3 次元カーネル密度推定法を用いてパス取得位置の確率密度関数を推定する．推定された密度関数を用いることで，任意の座標におけるパス取得の確率密度関数を計算でき，この値をそれぞれの条件でのパス成功/失敗に与える影響力と定義する．次に推定された 6 種の確率密度関数を用いて，各パスに対する全選手の影響力を算出する．算出した全選手の影響力に対してパス成功を 1・パス失敗を 0 としたラベルを付与し，ランダムフォレストを用いてパスの成功確率を予測する回帰モデルを学習した．この結果，予測値 0.5 以上を成功・0.5 未満を失敗とした際の予測精度は，パス成功が 88.7%，パス失敗が 91.1%となった．

4.2　スペースと選手の評価

学習したモデルを応用することでスペースや選手の評価を行う．まずスペースの評価として，フィールドを 1m 四方に分割し，分割した各区画に対してパスを出した際の成功確率を予測する．続いて全ての予測値をヒートマップとして表すことで，パスの成功確率に基づいた各チームの支配スペースを表すことができる(図 2)．また選手の評価として，各選手が実際に取ったパスと，そのパスに対する予測確率を比較する．予測確率が低い場所へのパスであったにもかかわらずパスを取ることに成功している選手は，敵のマークの外し方やトラップ技術など，パスを取る能力が優れていると考えパス取得時の予測確率の平均が低い選手を抽出した．この結果，例えば 2017 年 J リーグ MVP に選出された小林悠選手など実際に J リーグで活躍したフォワード選手が多く選出された．

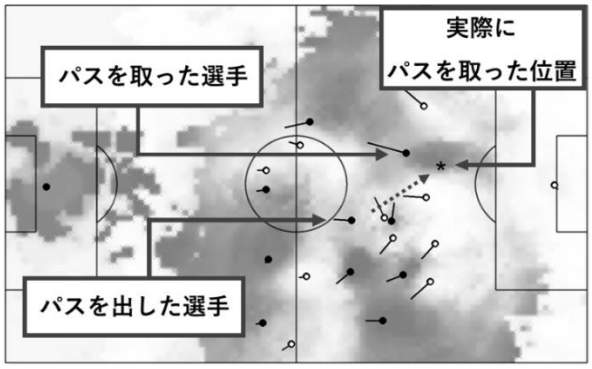

図 2 スペース評価の例

5　まとめ

本研究ではスポーツの試合から得られた実測データを対象とし，統計学や機械学習を用いて選手やチームの特徴抽出，及び可視化することを目的として分析を行った．まずフェンシングについては，試合の映像データを対象に突きと前後移動に着目して選手の特徴を表す指標を提案した．次にバスケットボールについては，行動履歴データを対象にタイムアウト前後のプレーに着目し，各チームが捉える試合の流れを可視化する手法を提案した．最後にサッカーについてはトラッキングデータを対象に，選手の位置関係や速度に着目してパスの成功確率を予測するモデルを提案し，パスの成功と失敗共に約 90%の精度で分類できることを示した．また提案モデルを応用することで，スペースや選手の評価が可能であることを示した．以上のように本研究では，3 種類のスポーツにおける異なるデータを対象とし，統計学や機械学習の観点から選手やチームに関する特徴を可視化する手法を示した．

謝辞

本研究を進めるにあたり，貴重なデータをご提供頂いたデータスタジアム株式会社様，並びに研究をご支援いただいた情報・システム研究機構 統計数理研究所 医療健康データ科学研究センター様に深く感謝いたします．

参考文献

[1]　成塚拓真，「対戦型集団スポーツの数理―サッカーを題材に―」，オペレーションズ・リサーチ 63, 10, 614-620, 2018.