

平成30年6月8日現在

機関番号：32641

研究種目：基盤研究(C)（一般）

研究期間：2015～2017

課題番号：15K00054

研究課題名（和文）大規模なスポーツデータに関する統計的モデリング

研究課題名（英文）Statistical Modeling for large-scale sports data

研究代表者

酒折 文武（Sakaori, Fumitake）

中央大学・理工学部・准教授

研究者番号：90386475

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：サッカーや野球などのスポーツでは選手やボールのトラッキングデータが収集され、選手の走行距離やスピード、ボールに関しない選手の動きなどを評価するようになってきた。スポーツ分野におけるデータ活用の重要性は高まってきている。本研究では、こうしたデータを活用し、プレイデータや統計情報、トラッキングデータ、さらには試合映像のようなスポーツ非構造化データを用いた統計分析法の確立と実践を目指した。その結果、スパースモデリング等を用いて野球の投手の肘故障の危険やその要因をトラッキングデータから明らかにする方法、サッカーのトラッキングデータからその時に起きているアクションの自動タグ付けの方法などを提案した。

研究成果の概要（英文）：Large-scale and complex data, e.g., tracking (trajectory) data of players and ball, are collected in several sports such as soccer and baseball. The importance of data and their use is increasing. In this study, we established some statistical methods and obtain useful knowledge by using trajectory data. Especially, we have developed a statistical analysis of medial collateral ligament injury using baseball tracking data, and automatic tagging methods for soccer tracking data.

研究分野：統計科学

キーワード：スポーツ統計科学

1. 研究開始当初の背景

セイバートリクス(野球統計学)に代表されるように、野球、バスケットボール、バレーボールなどのプロスポーツあるいはオリンピックスポーツを中心に、プレイ中の行動履歴データのリアルタイム/事後的な分析が活発に行われるようになってきた。これらは戦術判断や戦略評価、チーム編成のためにはなくてはならないものとなっている。さらに、サッカー、ラグビー、野球などデータ収集において先進的なスポーツでは、選手やボールの軌跡データ(トラッキングデータ)が収集されるようになり、行動履歴データでは測ることのできない、選手の走行距離やスピード、ボールに関しない選手の動きなどを評価するようになってきた。また、こうしたデータ活用が先進的ではないスポーツにおいても、試合映像の収集が進んでいる。

このように、スポーツ分野におけるデータ活用の重要性は高まってきているが、軌跡データは大規模かつ複雑であり、そのデータの情報量の割に行われている分析は貧弱であった。さらには、試合映像のような非構造化データについては、まったく分析に手がついていない。こうした大規模で複雑なデータや非構造化データについての新たな統計分析手法の確立が喫緊の課題であった。

2. 研究の目的

本研究では、スポーツにおける大規模で複雑なデータを活用することにより、以下を目指した。

(1) プレイデータや統計情報を活用した統計分析手法の確立とその実践

(2) 選手やボールの軌跡を記録したトラッキングデータを用いた統計分析手法の確立とその実践

(3) 試合映像のようなスポーツ非構造化データからの特徴抽出法の開発とその実践

3. 研究の方法

実際に、以下のような方法で研究を進めた。

(1) プレイデータの活用

若手選手と主力選手など、質の異なる空間の比較を行い、若手選手の将来予測を行う方法として、Cross-Domain Matching Correlation Analysis を用いる方法を検討した。

(2) トラッキングデータの活用

(2-a) サッカーにおけるトラッキングデータとプレイデータの2つのデータセットを結合し、2つのデータセットの時刻のずれによる不正確性も考慮に入れながら、選手間やボールの位置関係やアクションなどを用いて現在起きているプレイ状況を自動検出する

アルゴリズムの開発に取り組んだ。

(2-b) 野球における投手のリリースポイントの座標やばらつき(肘への負担を考慮)、速球の球速やボールの回転数などトラッキングデータから測定できる項目に加えて、登板間隔や1試合の投球数なども含め、投手の肘の故障に影響を与える要因についてのモデリングと実証分析に取り組んだ。

(3) 試合映像の活用

用いることのできる統計手法の検討、分析すべきスポーツの検討などを行った。

これら以外に関しても、おもにプレイデータを用いて、学生たちと共同で多くの分析を行った。

4. 研究成果

(1) Cross-Domain Matching Correlation Analysis を用い、野球における1球ごとの行動履歴データと年間成績データに基づいて若手打者と主力打者を一部の類似度を用いて同一空間へマッピングすることができた。ドメイン空間を観測変数の線形部分空間とすることで、活躍に必要な能力の同定を行うことも可能であり、さらにドメイン空間の非線形化により予測精度の向上が見込めることもわかった。これは統計数理研究所共同研究レポートに論文(査読なし)掲載されている。

(2a) 現段階では、トラッキングデータにおける多数の特徴量と、プレイデータも掛け合わせてのボールの動きや、そのときに起きているボールに関わるアクションなどを学習データとして、スパースロジスティック回帰とサポートベクターマシンといった基本的な生成モデルと予測モデルを用いて、局所的に起きているプレイの自動タグ付けをある程度の精度で行うことができることを明らかにした。これは統計数理研究所共同研究レポートに論文(査読なし)掲載されている。他のスポーツにおける隠れマルコフモデルなどを用いた研究事例が出てきたことから、今後はこうしたモデルを用いることを検討している。

(2b) 上述のデータを用いて、ロジスティック回帰モデルおよびスパースロジスティック回帰モデルを用いて、先発投手およびリリーフ投手に関する肘の故障に関わる要因を抽出することができ、一定の成果を上げることができた。これは統計数理に論文として掲載されている。

しかし、先発投手に関する分析結果が一部満足とはいえないところがあり、さらなる改善を目指した。具体的には、投手の疲労度や投手ごとの疲労時の傾向を掴むためのモデリングを行い、これらを表す特徴量の抽出を

おこなった。さらに、それらの特徴量を用いて、故障のリスク要因解明のためのモデルを改良した。これにより、判別器としての予測精度が大幅に向上するとともに、得られた結果の解釈も満足いくものとなった。この成果は9月の国際会議にて発表を行った他、現在論文投稿の準備中である。

(3) 研究期間においては検討に止まったが、期間終了後の平成30年以降に、フェンシングおよびバスケットボール、さらにはテニスに関する分析を行なっていく準備が整いつつある。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

[1] 酒折文武、圓城寺啓人、竹森悠渡、西塚真太郎、保科架風 (2017). 野球のトラッキングデータに基づいた肘内側側副靭帯損傷の要因解析. 統計数理、65巻2号、201-205.

[学会発表](計17件)

[1] 辻響実、片山元貴、飯田歩、上田朋矢、酒折文武 (2017). 次なる優勝請負人は? 見つけ出せ、MLBの秘宝!!、シンポジウム「スポーツアナリティクスと統計科学」

[2] 辻響実、片山元貴、飯田歩、上田朋矢、酒折文武 (2017). 次なる優勝請負人は? 見つけ出せ、MLBの秘宝!!、シンポジウム「スポーツアナリティクスと統計科学」

[3] Fumitake Sakaori、Ibuki Hoshina (2017). Risk factors and predictive modeling of the UCL injury using PITCHf/x data in MLB. New England Symposium on Statistics in Sports

[4] 酒折文武、保科架風 (2017). PITCHf/xデータに基づくMLB投手のUCL損傷要因の考察、2017年度統計関連学会連合大会

[5] 圓城寺啓人、竹森悠渡、西塚真太郎、保科架風、酒折文武 (2017). 160km/hよりも打ちにくい130km/hー甘いコースでも打たれないファストボールのモデリングー、シンポジウム「スポーツアナリティクスと統計科学」

[6] 石原渚、大屋拓磨、羽山美優、望月優志、細田幸希、保科架風、酒折文武 (2017). ハリルも唸る、次なる代表選手はだ!〜デュエルの自動検出と活用〜、シンポジウム「スポーツアナリティクスと統計科学」

[7] 入江敦子、柴田祐大、酒折文武 (2017). 8年ぶりのWBC王座へ!2017年WBC日本最強ラインアップの構築〜世界大会で活躍でき

る打者の特徴とは〜、シンポジウム「スポーツアナリティクスと統計科学」

[8] 圓城寺啓人、竹森悠渡、西塚真太郎、保科架風、酒折文武 (2016). 160km/hよりも打ちにくい130km/hー甘いコースでも打たれないファストボールのモデリングー、シンポジウム「スポーツアナリティクスと統計科学」

[9] 石原渚、大屋拓磨、羽山美優、望月優志、細田幸希、保科架風、酒折文武 (2016). ハリルも唸る、次なる代表選手はだ!〜デュエルの自動検出と活用〜、シンポジウム「スポーツアナリティクスと統計科学」

[10] 入江敦子、柴田祐大、酒折文武 (2016). 8年ぶりのWBC王座へ!2017年WBC日本最強ラインアップの構築〜世界大会で活躍できる打者の特徴とは〜、シンポジウム「スポーツアナリティクスと統計科学」

[11] 酒折文武 (2016). サッカーのトラッキングデータに関する統計的モデリング、2016年度統計関連学会連合大会

[12] 君島亮、齋藤秀哉、竹森悠渡、中嶋雅彦、西塚真太郎、酒折文武 (2016). 大谷翔平はどんなバッターになるのか?ー最新手法CDMCAを用いた若手選手の発掘ー第5回スポーツデータ解析コンペティション受賞者講演会、

[13] 井口雅章、高橋達也、細田幸希、望月優志、酒折文武 (2016). チェンジアップで柳田を料理〜アソシエーション分析を用いて〜、第5回スポーツデータ解析コンペティション受賞者講演会

[14] 君島亮、齋藤秀哉、竹森悠渡、中嶋雅彦、西塚真太郎、酒折文武 (2015). 大谷翔平はどんなバッターになるのか?ー最新手法CDMCAを用いた若手選手の発掘ー、第5回スポーツデータ解析コンペティション発表審査会

[15] 井口雅章、高橋達也、細田幸希、望月優志、酒折文武 (2015). チェンジアップで柳田を料理〜アソシエーション分析を用いて〜、第5回スポーツデータ解析コンペティション発表審査会

[16] 石原渚、圓城寺啓人、大屋拓磨、酒折文武 (2015). 主婦が料理してても手を止める!? 魅力的な野球中継〜ロジスティック回帰を使った得点率の推測〜、第5回スポーツデータ解析コンペティション発表審査会

[17] 酒折文武 (2015). サッカーのトラッキ

ングデータに関する統計的分析の可能性について、2015年度統計関連学会連合大会

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕

論文特集号1編の編集を行った。
統計数理第65巻2号特集「スポーツ統計科学の新たな挑戦」

本研究成果に関わる書籍「統計学 One Point スポーツアナリティクス (仮称)」(共立出版)の発行を準備している。

統計数理研究所共同研究レポート(査読なし論文)として複数の原稿が掲載されている。

本研究で得られた知見から、日経新聞 2018年5月31日~6月7日に記事を連載した。

研究成果の一部(野球における故障の要因解析)が、2018年4月中旬の山陽新聞、南日本新聞、新潟日報、秋田さきがけ、下野新聞、千葉日報、日本海新聞、岐阜新聞、京都新聞、信濃毎日新聞、熊本日日新聞、大分合同新聞に取り上げられた。

ホームページ情報

研究成果の一部は以下のウェブサイトに掲載されている。

<http://www.math.chuo-u.ac.jp/~sakaori/>
<http://estat.sci.kagoshima-u.ac.jp/sports/>

6. 研究組織

(1) 研究代表者

酒折 文武 (SAKAORI, Fumitake)

中央大学・理工学部・准教授
研究者番号：90386475

(2) 研究分担者

山口 和範 (YAMAGUCHI, Kazunori)
立教大学・経営学部・教授
研究者番号：60230348

渡辺 美智子 (WATANABE, Michiko)
慶應義塾大学・健康マネジメント研究科・教授
研究者番号：50150397

(3) 連携研究者

田村 義保 (TAMURA, Yoshiyasu)
統計数理研究所・統計思考院・特任教授
研究者番号：60150033

竹内 光悦 (TAKEUCHI, Akinobu)
実践女子大学・人間社会学部・教授
研究者番号：60339596