

フットボールアナリティクスのサッカーゲームデータ分析研究の動向

西嶋尚彦*・松岡弘樹*・安藤 梢*

Trends in soccer game data analysis research in football analytics

NISHIJIMA Takahiko*, MATSUOKA Hiroki* and ANDO Kozue*

Abstract

Football analytics is a sports analytics specialized in soccer and has developed remarkably in recent years, especially in Europe. Football analytics is the study of soccer data analysis based on the development of new data technologies that combine satellite GPS technology, camera technology, and computer science in soccer stadiums and club training facilities. Recent trends in soccer game data analysis research in football analytics were reviewed using the review papers. Analysis techniques for soccer game data can be broadly classified into two categories of notational analysis and motion analysis. Notational analysis is a method for objectively recording play events such as passing, dribbling, and shooting in soccer games. Since 2010, based on studies that applied machine learning to ball event data and tracking data measured from soccer games using data technology, feature items such as player and ball movement trajectories derived from tracking data, defensive indicators consisting of the positions of defensive players, and image data plotting player location information were developed. Tracking data measured from soccer games are essential for football analytics research with motion analysis. Game stat data were compiled from soccer game ball event data measured from notational analysis. In a study using game stats data, measures of soccer game wins, losses, playing success, game performance, and style of play were developed. Scale construction methods with multivariate statistical analysis for testing physical fitness and motor ability and for testing exercise lifestyle behavior questionnaires were applied to develop items measuring soccer skills from soccer game performance.

Keywords: notational analysis, motion analysis, machine learning, item development, tracking data

1. フットボールアナリティクスの研究動向

フットボールアナリティクス (Football analytics, フットボール分析学) はサッカーに専門的なスポーツアナリティクス (Sports analytics, スポーツ分析学) であり、近年において欧州を中心に顕著に発展している。サッカースタジアムやクラブのトレーニング施設における人工衛星による GPS 技術、カメラ撮影技術、コンピュータサイエンスの結合による新しいデータテクノロジーの発展を基盤とするサッ

カー分析学である。総説論文 (Review paper) を手掛かりとして、近年のフットボールアナリティクスに関するサッカーゲームデータ分析研究の動向を検討した。表 1 は、2008 年以降におけるサッカーゲームデータ分析研究に関する総説論文 (レビュー論文、システムティックレビュー論文、スコアリングレビュー論文) を示している。

Carling et al.³⁾ のエリートサッカーにおけるモーションアナリシス (Motion analysis) の役割に関す

* 筑波大学体育系
Faculty of Health and Sport Sciences, University of Tsukuba

るレビューでは、エリートサッカーを対象とする様々なモーションアナリシスシステムに関する研究者と実践者の双方の認識を高め、構築された知識体系から実践的な役割が明示され、さらなる探求が必要な領域であることを指摘した。

Mackenzie and Cushion²¹⁾ のサッカーのパフォーマンス分析に関する批判的レビューと今後の研究への示唆では、サッカーのゲームパフォーマンス分析は予測変数やパフォーマンス制御変数の研究に重点を置き過ぎていることを明らかにした。アスリートとコーチの学習に関する調査研究の発展のためのパフォーマンス分析情報と連携する別のアプローチを提案した。今後のサッカーのパフォーマンス分析研究では、応用分野におけるパフォーマンス分析の実施とアスリートの学習に影響を与える社会・文化的影響に注目する必要があることを指摘した。

Sarmento et al.²⁸⁾ のサッカーの試合分析 (Match analysis) に関するシステムティックレビューでは、レビューされた研究の限界を統括し、研究された変数の包括的な運用定義を提供すること、活動や参加者の標準化されたカテゴリと説明を使用すること、サッカーのパフォーマンスの分析に状況や相互作用のコンテキストを統合することを、今後の研究で検討する必要があることを指摘した。

サッカーのデータテクノロジーの発展により、2016年以降の総説論文のタイトルには、Big data、Tracking data、Positional data、Machine learningなどの用語が使用されていた。

Rein and Memmert²⁶⁾ のエリートサッカーにおけるビッグデータと戦術分析におけるスポーツ科学の今後の課題と可能性に関するレビューでは、エリートサッカーの戦術分析における問題を検討し、エリートサッカー研究にビッグデータ技術を導入するための技術スタックが必要であることを提案した。この提案は、データサイズの増大が進展している他のスポーツ科学領域におけるガイドラインとなるものであった。

Memmert and Rein²⁴⁾ は、エリートサッカーにおける試合分析、ビッグデータ、戦術の最近の傾向についてレビューし、科学的根拠が既に明らかであるエリートサッカーの新しいパフォーマンス指標を紹介した。パフォーマンス分析はビッグデータ分野との関連がますます強まることを予想し、ビッグデータと試合分析を結びつけるモデルが、将来のトレーニング実践を想定する結果を示すことを実証した。

Herold et al.¹⁰⁾ のレビューでは、男子プロサッカーにおける機械学習について攻撃的なプレーを向上させるための現在の応用と将来の方向性について検討

し、戦術的な知識やパフォーマンスを向上させるための機械学習に関する研究はほとんどなく、レビュー論文もないことを指摘した。

2020年以降では、Collective tactical behaviors、Tactical performance、Defensive playなどの用語がタイトルに使用されていた。サッカーゲームデータ分析研究は、ゲーム単位の分析から、守備戦術、攻撃戦術、攻守の切替戦術に対応するプレー単位の分析をするフットボールアナリティクスへ発展した。

Low et al.¹⁹⁾ は、位置情報を用いたサッカーの集団的戦術行動に関するシステムティックレビューにおいて、本分野は科学的進歩が小さいが、確実な伸展を示している成長中の研究分野であることを指摘した。

Goes et al.⁹⁾ のプロサッカーにおける戦術的なパフォーマンス分析をサポートするためのビッグデータの可能性の解放に関するシステムティックレビューでは、近年のスポーツ科学とコンピュータサイエンスの領域において、位置追跡データを用いた戦術的な行動に関する研究が大きく発展しているものの、スポーツ科学は理論や実用的な意味合いに重点を置き、コンピュータサイエンスは技術の開発に重点を置いており、両者の研究成果は明確に異なっていることを明らかにし、スポーツ科学とコンピュータサイエンスのコラボレーションにおける以下のような要点を指摘した。

- 1) サッカーの戦術研究において、特にポジショントラッキングのデータを用いる場合には、スポーツ科学とコンピュータサイエンスのコラボレーションにかなりの機会が存在する。
- 2) スポーツ科学とコンピュータサイエンスのコラボレーションは、より強い対話によって、循環的なコラボレーションを実現することが可能である。
- 3) サッカーの戦術分析において、スポーツ科学とコンピュータサイエンスの専門知識を組み合わせるための基礎となるフレームワークを提案する。

Forcher et al.²²⁾ の選手追跡データを用いたプロサッカーのディフェンスプレー分析に関するスコピングレビューでは、フットボールアナリティクスにおけるトラッキングデータを用いたサッカーの守備的プレー分析研究の成果の特徴を検討し、以下のような要点を明らかにした。

- 1) トラッキングデータを用いたサッカーの守備的プレーの分析アプローチは、非常に異質である。
- 2) ディフェンスプレーを分析する最も有望なアプローチは、ディフェンスプレッシャーの分析 (個

人レベル)と、コンパクトオーガニゼーションの定量化(グループレベル)である。

- 3) 成功するディフェンスプレーの特徴は、高いディフェンスプレッシャー(個人レベル)、チーム間・チーム内の同調とバランスのとれたディフェンス(グループレベル)、組織の縮小(チームレベル)である。

データテクノロジーを使用してサッカーゲームから計測されたボールイベントデータとトラッキングデータに機械学習を適用することにより、得点やシュートなどの価値の高いプレーに貢献するサッカー選手の評価方法の開発研究が進展していくことが予想される。フットボールアナリティクスの目的は、サッカー選手の発掘と育成だけではなく、マーケティングの観点から顕在化されていない選手の価値を適切に評価する項目を開発することである。データテクノロジーを基盤とするサッカーゲームデータ計測の発展を考慮すると、フットボールアナリティクス研究では、サッカーゲームのビッグデータから新しい実践知を得ること、サッカーの専門家もつ暗黙知の計量から新しい形式知を得ることが期待される。

ビジネスが先行する現状のサッカーゲームデータ分析研究における課題は、研究者がサッカーゲームデータのビッグデータを入手が困難なことである。

このような重要な課題を解決するために、Pappalardo et al.²⁵⁾は、2017-18シーズンにおけるPremier League(イングランド)、La Liga(スペイン)、Ligue 1(フランス)、Bundesliga(ドイツ)、Serie A(イタリア)とFIFA World Cup 2018、UEFA Euro Cup 2016のイベントデータを、CC BY 4.0 Licenseの下、figshare(https://figshare.com/collections/Soccer_match_event_dataset/4415000/5)で公開した。このサッカーゲームのビッグデータは、ヨーロッパ5大リーグ(イングランド、スペイン、フランス、ドイツ、イタリア)と国際大会の1,941試合、3,251,294イベント、4,299選手のデータが記録されたビッグデータである。選手のトラッキングデータは含まれていないものの、イベントデータにはボールの位置情報が記録されているため、分析価値の高いデータセットである。

2. サッカーゲームデータ計測技術の動向

フットボールアナリティクス研究の動向は、サッカーゲームデータ分析研究における計測技術の発展と計測されたデータに適合した分析技術の発展に関する歴史、すなわち科学技術史の観点から見る事ができる。

表2は、サッカーゲームデータの計測技術、データ、研究技術、研究内容の動向を示している。サッ

Table 1. Trends in review articles on soccer game data analysis research.

| Year | Author | Title | Type | Journal |
|------|---------------------|--|-------------------|--|
| 2008 | Carling et al | The role of motion analysis in elite soccer | Review | Sports Medicine |
| 2013 | Mackenzie & Cushion | Performance analysis in football: a critical review and implications for future research | Critical review | Journal of Sports Sciences |
| 2014 | Sarmiento et al | Match analysis in football: a systematic review | Systematic review | Journal of Sports Sciences |
| 2016 | Rein & Memmert | Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science | Review | SpringerPlus |
| 2018 | Memmert & Rein | Match analysis, big data and tactics: current trends in elite soccer | Review | Deutsche Zeitschrift für Sportmedizin |
| 2019 | Herold et al | Machine learning in men's professional football: Current applications and future directions for improving attacking play | Review | International Journal of Sports Science & Coaching |
| 2020 | Low et al | A systematic review of collective tactical behaviours in football using positional data | Review | Sports Medicine |
| 2021 | Goes et al | Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review | Systematic review | European Journal of Sport Science |
| 2022 | Forcher et al | The use of player tracking data to analyze defensive play in professional soccer - A scoping review | Scoping review | International Journal of Sports Science & Coaching |

Table 2. Trends in measurement techniques, data, research techniques, and research content for soccer game data.

| Year | Measurement techniques | Data | Research techniques | Research contents |
|-----------|---|---|---|---|
| 1970～ | Hand-notational analysis | Game stats | Descriptive statistical analysis | Counting items for offensive play |
| 1990～2000 | Computer-notational Analysis | Game stats + Game performance | Descriptive, + Multivariate statistical analysis | Development of game performance items, key performance indicators (KPI) |
| 2000～2010 | Motion analysis, Computer-notational Analysis | Game stats + Game performance + Tracking | | Development of team tactical action items using tracking data |
| 2015～ | | Game stats + Game performance + Tracking + Game event | Descriptive, + Multivariate statistical analysis + Machine learning | From big data of soccer games development of feature items for attacking Plays, for transition plays, for defensive plays |

カーゲームデータの分析技術は、ノーテーションアナリシス（表記分析、Notational Analysis）とモーションアナリシス（モーション分析、Motion Analysis）に大きく分類される。ノーテーションアナリシスは、サッカーゲームにおけるパス、ドリブル、シュートなどのプレイイベントを客観的に記録する手法である。1970年代から1990年代では、記述的な分析手法であるハンドノーテーションアナリシス（手記表記分析、Hand Notational Analysis）であった。大型計算機システムの普及とパーソナルコンピュータ（PC）用SOのWindows95の登場により、1990年代ではコンピュータノーテーションアナリシス（コンピュータ表記分析、Computer Notational Analysis）へ発展した。2000年代のコンピュータ技術の発展に伴い、コンピュータノーテーションアナリシスでは、短時間で大量のゲームパフォーマンスデータとゲームイベントデータが計測された。

一方、モーションアナリシスは、サッカーゲームにおける選手個人の行動や移動を客観的に記録する手法である。2005年頃からは、Global Positioning System（GPS、全地球測位システム）に代表される衛星測位システムや、複数台の高性能カメラを使用するコンピュータビデオトラッキングシステム（Computerized Video Tracking Systems）などのデータテクノロジーの発展により、サッカーゲームにおける選手の位置を追跡し、選手の位置情報を時系列データとして取得する計測技術が発展した。

近年では、ノーテーションアナリシスとモーションアナリシスの計測技術が発展し、サッカーゲームから時系列のビッグデータが計測され、デー

タの特徴から以下のような4種類に分類される。第1はゲームスタッツデータであり、試合単位のシュート回数やパス回数が記録され、集計された計数データである。第2はゲームパフォーマンスデータであり、ポゼッションプレーにおけるパスの質のような専門家の視点からゲームパフォーマンスを計測したデータである。第3はボールイベントデータであり、サッカーゲーム1試合を通じたボールイベントとイベント時のピッチ上の位置情報（ x と y の2次元座標値）が時系列に記録されたビッグデータである。そして、第4はトラッキングデータでありGPSや高性能カメラからサッカー選手の位置情報（ x と y の2次元座標値）が時系列に計測されたビッグデータである。

ゲームスタッツデータ、ゲームパフォーマンスデータ、ゲームイベントデータ、トラッキングデータの順にデータ量が増大する。ゲームイベントデータでは1試合で約2,000イベントのデータが取得され、トラッキングデータでは1試合で約150万件のレコード（1試合で約7万件のレコード×選手22人）が取得されるビッグデータである。これらの2種類のデータを加工して、組換え、構造方程式モデリング（Structural Equation Modeling: SEM）による検証的因子分析（Confirmatory Factor Analysis: CFA）や達成度評価（目標規準準拠測定、Criterion-referenced Measurement）のテスト理論である項目反応理論（Item Response Theory: IRT）分析などの多変量統計解析に加えて、最先端のデータサイエンス技術である機械学習を適用することにより、サッカーのゲーム分析とプレー分析のためのKey Performance Index（KPI）項目、ゲームパフォーマンス項目、プレー

予測・識別モデルのための特徴量項目などの新しい測定項目が開発されてきている。そして、これらのサッカーのスポーツアナリティクス (Sports Analytics) 分野での研究成果からフットボールアナリティクス (Football Analytics) が発展している。

従来のゲームパフォーマンス分析では、選手とチームの評価やパフォーマンスを改善するトレーニングプログラム設計のための補助資料として、監督やコーチの指導者がノーテーションアナリシスの結果を活用していた (Carling et al. 2005; Wright et al. 2012)。これに比較して、ゲームデータ分析の知見は、近年のデータテクノロジーの発展に伴うサッカーゲームから計測されるビッグデータから分析されており、データ主導型のトレーニングマネジメント、ゲームマネジメント、クラブマネジメント、リーグマネジメントへと活用範囲が拡大し、サッカーコーチングの指導者と選手の利用に限らず、サッカーリーグやサッカークラブの経営者が活用している。

フットボールアナリティクスにおけるサッカーゲームデータ分析研究における学術的ニーズは、データテクノロジーの発展により計測される新しいサッカーゲームデータから、未だ解明されていないサッカーゲーム特性を測定する新しい項目を明らかにすることである。

フットボールアナリティクス研究におけるサッカーゲームデータ分析は、計測されたデータの発展経緯から以下の4種類に大別することができる。第1はボールイベントデータとトラッキングデータに機械学習を適用した研究である。第2はモーションアナリシスから計測されるトラッキングデータを用いた研究である。第3はノーテーションアナリシスから計測されるゲームスタッツデータを用いた研究である。そして、第4はゲームパフォーマンスデータやサッカーゲームのビッグデータを用いたサッカーゲームの尺度構成 (項目開発) 研究である。

3. 機械学習によるフットボールアナリティクス研究の動向

サッカーゲームデータ計測におけるデータテクノロジーの発展を考慮すると、機械学習 (マシンラーニング、Machine learning) 技術の適用はフットボールアナリティクスの今後の発展に必要不可欠である。特に、深層学習 (ディープラーニング、Deep learning) アルゴリズムの適用による画像認識技術開発が進展することが予想される。2010年以降では、データテクノロジーを使用してサッカーゲームから計測されたボールイベントデータとトラッキン

グデータに機械学習を適用した研究により、トラッキングデータから得られた選手とボールの移動軌跡、守備選手の位置から構成された守備指標、選手の位置情報をプロットした画像データのような特徴量項目が開発された。

Lucey et al.²⁰⁾ は、サッカーゲームにおける選手の位置情報から計測される戦術的な特徴量項目に機械学習を適用して、サッカーゲームにおける得点期待値を推定した。特徴量項目は、Match context (オープンプレー、カウンターアタック、セットプレー)、Location (シュート位置とゴールの距離)、Defender proximity (シュート角度内の守備選手の距離)、Defensive formation/structure (守備ライン間の距離)、Attacking features (ロングパス、クロス、ドリブル) であった。ロジスティック回帰モデルを適用して、得点期待値を推定した。Match context、Location、Defending features、Attacking features を独立変数としたロジスティック回帰モデルの誤差は0.143であり、最も良好なモデルであった。シュートの位置やプレーの種類だけでなく、守備と攻撃におけるチーム戦略的な特徴量項目を開発することにより、モデル精度が向上することが示唆された。標本は、20チームの353試合におけるボールトラッキングデータとプレーヤートラッキングデータから抽出されたシュート直前10秒間における9,732シュートプレーデータであった。

Brooks et al.²¹⁾ は、サッカーゲームにおけるパスのイベントデータから、チームのパススタイルを分類する方法とシュートに繋がるパスの特徴を分析する方法を提案した。パス項目は、パス開始位置 (Origin of the pass)、パス到達位置 (Destination of the pass)、パス結果 (Outcome of the pass; successful or not)、パスの出し手 (Player attempted the pass)、パスの受け手 (Intended recipient of the pass)、時間 (Time of the pass)、チーム (Team)、試合 (Game) であった。チームのパススタイル分類データからチーム毎のパスヒートマップを作成し、ヒートマップを特徴量としたK近傍法 (KNN) を適用して、チームのパススタイルを分類した。分類精度は87%であった。特徴量項目は、ポゼッション毎、ピッチゾーン毎 (18ゾーン) に、Origin grid、Destination grid、Origin-destination grid であった。サポートベクターマシーン (SVM) を適用して、パスデータからのシュート予測モデルを開発した。その後、Origin grid、Destination grid、Origin-destination grid の特徴量の重みを用いて、選手をランク付けし (Pass shot value)、評価した。シュート予測モデルの精度はAUROC = 0.785であり、Pass shot value の高い選手はクリス

ティアノ・ロナウドとリオネル・メッシであった。標本は、2012-2013年スペインリーグ20チームのパスデータであり、チーム平均18,000本であった。

Wagenaar et al.³⁴⁾は、深層学習の畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を適用して、サッカーゲームの得点機会の予測モデルを構築した。トラッキングデータの選手の位置を基に、256 × 256ピクセルの画像データを作成した。CNNはGoogleNet、3-Layer Convolutional Neural Networkを適用し、ベースラインモデルとしてK近傍法 (KNN) を適用した。GoogleNetの予測精度は67.1%であり、最も高い結果であった (KNN=57.3%)。分析結果は、画像分類の深層学習が得点機会の予測に活用できることを示した。標本は、ドイツブンデスリーガの29試合 (2009年) のトラッキングデータであった。

Link and Hoernig¹⁸⁾は、サッカーゲームのトラッキングデータに機械学習を適用し、チームと選手個人のポジションプレーの自動識別モデルを開発した。チーム単位だけでなく、個人単位のポジションプレーを分析し、選手がボールに対して行動するプレー局面を示すIndividual ball actionを特定した。特徴量項目は、Individual ball actionの時間 (Duration)、平均ボール速度 (Average ball velocity)、平均ボール加速度 (Average ball acceleration)、ボール速度の分散 (Variance of ball velocity)、ボール加速度の分散 (Variance of ball acceleration)、選手とボールの平均距離 (Average distance between the ball and the player in possession of the ball) であった。ベイジアンネットワーク (Bayesian network) を適用して、ボールコントロールの程度を示すIndividual ball controlを分類した。ポジション毎のIndividual ball controlの時間を推定することが可能であり、センターフォワード (Central FW) のポジション時間はゴールキーパー (GK) とセンターディフェンダー (Central DF) と比較して短いことを示した。標本は、2013-2014年ドイツブンデスリーガの60試合であった。

Hobbs et al.¹¹⁾は、機械学習を適用して、サッカーゲームにおけるカウンターアタック (Counter attack) を客観的に自動的に分類する方法を提案した。90分間のサッカーの試合を10秒区間に分割し、分割区間における選手のトラッキングデータに対して階層的クラスタ分析 (Hierarchical cluster analysis) を適用して、選手の移動パターンの特徴からプレーを9種類 (218分類) に分類した。切替プレーは14分類であった。切替プレーを対象とする守備指標のDefensive disorderと攻撃指標の

Attacking threatを開発した。Defense disorderは、切替プレー分類における最適守備選手のフォーメーションに対する実際の守備選手のフォーメーションの差異を点数化した指標であった。Attacking threatは得点期待値を示す指標であった。2項ロジスティック回帰分析を適用して、次の10秒間におけるシュート確率が推定された。Attacking threatから脅威のある攻撃と脅威のない攻撃に識別して、切替プレーの中からカウンタープレーを分類した。切替プレーの中から自動的に分類されたカウンターアタックのシュート確率は35.1%であり、それ以外の攻撃 (10.6%) と比較して高かった。同様に、カウンターアタックのゴール期待値は12.4%であり、それ以外の攻撃 (8.0%) と比較して高かった。また、カウンターアタックのシュートまでの平均時間は12.9秒であり、それ以外の攻撃の平均時間 (28.3秒) と比較して短かった。これらの結果から、サッカーゲームにおける10秒区間の選手のトラッキングデータに機械学習を適用することで、プレーの自動分類が可能となり、プレー分析の発展に貢献できることを示した。

Goes et al.⁸⁾は、サッカーゲームから計測されたトラッキングデータに機械学習を適用して、パス効果 (pass effectiveness) を定量化する方法を提案した。守備崩壊性 (Defensive disruptiveness) の指標として、守備選手とチームの位置情報から守備選手の動き指標 (I-Mov) と守備組織の崩壊指標 (D-Def) を算出した。守備組織の崩壊指標は、チームのx、yの相対的移動距離、ライン毎 (at, mf, df) の相対的移動距離、面積の変化、面積の拡大の特徴量項目から構成された。主成分分析を適用して3主成分を抽出し、3主成分得点の絶対値の総和を守備組織の崩壊指標得点とした。重回帰分析を適用して、守備選手の動き指標と守備組織の崩壊指標に影響するパス効果要因は、パス速度とパス距離 (速いパスと長いパス) であることが明らかとなった。標本は、2017-2018年オランダサッカーリーグの18試合から計測された10,481パス (全16,943パス) であった。

Decroos et al.⁴⁾は、サッカーゲームにおけるボールイベントのビッグデータに機械学習を適用して、サッカーゲームにおける選手の貢献度を評価する新しい指標を開発した。複数のスポーツ分析会社が記録しているボールイベントデータに共通する特徴量項目から構成されるデータセット (Soccer Player Action Description Language: SPADL) を構築した。SPADLに機械学習を適用して、各アクションの価値を定量化するフレームワークを検討し、各アクションが得点に繋がる確率と失点に繋がる確率から

VAEP (Valuing Actions by Estimating Probability) を算出した。各選手のVAEPの総和を時間で基準化し、選手のアクション貢献度を定量化した。VAEPによる選手の貢献度評価では、得点数やアシスト数からは評価が困難であった市場価値が高い選手の評価が可能となった。このように、複数のデータ分析会社が収集しているボールイベントデータから、機械学習に適用可能な共通のフォーマットを保持したデータセットであるSPADLを構築し、個々のアクションの価値を算出する分析フレームワークを開発することで、サッカーゲームにおける選手の貢献度を定量化した。標本は、イギリス、スペイン、ドイツ、イタリア、フランス、オランダ、ベルギーの2012/2013から2017/2018シーズンまでの11,565試合のボールイベントデータ（イベントストリームデータ）であった。

Jo et al.¹⁴⁾ は、2011年Jリーグ686試合のサッカーのボールイベントデータから構成したサッカーの攻撃プレーを測定する45項目に、Decision Tree、Random Forest、Gradient Boosting Decision Treeの機械学習モデルを適用し、シュート予測モデルを開発した。攻撃プレー23項目にGradient Boosting Decision Treeを適用したシュート予測モデルの精度が最も良好であった。また、機械学習モデルにおける特徴量の重要度から、Vital Areaの重要度が最も高いことを明らかにした。

4. トラッキングデータによるフットボールアナリティクス研究の動向

データテクノロジーの発展により、サッカーゲームから計測されたトラッキングデータは、モーションアナリシスによるフットボールアナリティクス研究に必要不可欠である。サッカーゲームのトラッキングデータから得られる位置情報を用いたモーションアナリシス研究により、チームの戦術行動(Tactical behavior)を測定する項目が開発された。トラッキングデータから、チームの中心位置(Center of a team、またはCentroid)、2チームの中心位置間の距離、チームの縦幅(Length)、横幅(Width)、面積(Surface are)に代表される測定項目が開発された。

Frencken et al.⁷⁾ は、サッカーゲームの動的変動を測定するチームの中心位置項目とサッカーゲームのイベントの関係を分析し、攻撃チームが攻撃方法を探している時の守備チームの戦術行動を定量化することが可能であり、チーム戦術を測定する新しい指標項目として活用できることを示した。チームの中心位置項目は、攻撃チームと守備チームの中心位置

間の縦距離(Longitudinal inter-team distance)と横距離(Lateral inter-team distance)であった。2項目の平均値±3標準偏差を超える変動があった場合、クリティカルイベントとした。クリティカルイベントが発生する局面は、得点と得点機会、攻撃チームのパス直後の守備チームのプレッシング、ボールアウトの3局面であった。縦距離(Longitudinal distance)の変動における93%のクリティカルイベントは縦パス時点であり、すべての選手が前方向または後ろ方向に移動していた。横距離(Lateral distance)の変動における87%のクリティカルイベントは横パス時点であり、守備選手が正しくポジションを横方向に移動していた。また、得点と得点機会では、攻撃チームと守備チームの中心位置間の距離(inter-team distance)に大きな変動が確認された。標本は2008-2009 UEFA Champions Leagueの準々決勝の1試合であった。

Sampaio and Macas²⁷⁾ は、サッカーゲームにおけるGPSを使用して計測されたトラッキングデータから測定される変数に近似エントロピーを適用して、サッカーの戦術行動(Tactical behavior)が測定できることを示した。被験者は大学生12名であり、Pre-Postデザインの実験を実施した。トレーニング期間は13週間であり、週2時間のサッカーの授業を実施した。PreテストとPostテストでは12分間の5対5のミニゲームを実施し、ミニゲームにおけるチームの中心位置、チーム中心座標、選手個人の座標値の距離、チームの中心位置から最も遠くにいる選手の距離、チームの中心位置から最も近くにいる選手の距離の近似エントロピー(Approximate Entropy: ApEn)を算出した。近似エントロピーは時系列データの定常性を示す指標であり、規則正しいデータの場合は0に近似し、ばらつきのある予測できないデータの場合は大きな値になる。ほぼ全ての項目の近似エントロピーは、PreテストとPostテスト間で低下した。つまり、Postテストでは選手の動きが安定し、サッカーの専門性が増大していることが示唆された。

Folgado et al.⁹⁾ は、年齢とミニゲーム形式の違いにより、チームの縦横比(Lpwratio)と攻守両チームの中心位置間の距離(Centroid distance)は変化し、ミニゲームにおける戦術トレーニングの意味を理解することに役立つことを示した。チームの縦横比(Lpwratio: Length ÷ Width)と攻守両チームの中心位置間の距離(Centroid distance)の2項目を従属変数とし、年齢(U9、U11、U13)とミニゲーム形式(3v3、4v4)を独立変数とする2要因分散分析を実施した。チームの縦横比は年齢要因の影響を受け、

若い年齢の方がチームの縦横比が大きかった。同様に、チームの縦横比の変化量は高い年齢の方が小さく、高い年齢の方が高い集団的な戦術行動であることを明らかにした。攻守両チームの中心位置間の距離では、3v3のミニゲームでは高い年齢の方が長い距離を示し、4v4のミニゲームではすべての年齢で近似した値を示した。U9、U11、U13の各年齢10人のサッカー選手がミニゲーム実験に参加した。

5. ゲームスタッツデータによるフットボールアナリティクス研究の動向

ノーテーションアナリシスから計測されたサッカーゲームのボールイベントデータを集計したゲームスタッツデータを用いた研究では、サッカーゲームの勝敗、プレーの成功、ゲームパフォーマンス、プレースタイルに関する測定項目が開発された。サッカーゲームのプレー成功に係る要因を分析した研究における測定項目は、パス本数、タックル、ロングパス、ドリブルなどに代表され、ボールイベントデータから構成された。

Barreira et al.¹⁾ は、攻撃結果とボール奪取 (Ball recovery) との関係进行分析し、攻撃結果に係る守備イベントを明らかにした。タックルによるボール奪取と得点の関係、パス後の守備行動 (Defensive behavior followed by a pass) と枠内シュートの関係を明らかにした。標本は、FIFA World Cup 2010の準決勝に進出した4チームの24試合における1,619守備プレーデータであった。

Kempe et al.¹⁵⁾ は、サッカーゲームの攻撃アクション項目から戦術行動 (tactical behavior) を評価するための新しい指標を開発した。パスに関連する項目から Index of Game Control (IGC) を開発した。IGCとゲームスピードや移動距離の項目を用いて、プレースタイル (style of play) を評価する Index of Offensive Behavior (IOB) を開発した。IOBはプレースタイルの特徴 (ポジションプレー vs ダイレクトプレー) を測定する妥当な項目であった。標本は、Bundesliga (2009-2010、2010-2011) と FIFA World Cup 2010の全676試合の各ポジションプレー (team possession) であった。

Hughes and Lovell¹²⁾ は、攻撃への切替 (Offensive transition) プレーの特徴を分析した。オフENSIVEゾーン (Offensive zone) での切替プレーとタックルでの切替プレーが、得点や得点機会に係ることを明らかにした。また、攻撃への切替直後の最初の2プレーが重要であり、ロングパス (Long pass)、ドリブル (Dribble)、ラン (Run) などのボールを前に運ぶイベントが攻撃を成功するために最適な行

動であることを明らかにした。標本は、2014-2015年の欧州チャンピオンズリーグのトーナメントステージの29試合における3,077プレーであった。

6. 測定項目開発によるフットボールアナリティクス研究の動向

1980年代のサッカー研究では、サッカー技能を測定するスキルテスト (Skill test) が開発された。サッカーのドリブル技能を測定する8の字ドリブルテストに代表される実技テスト (Performance test) が開発された。再テスト法 (Test-retest method) によるテストの信頼性ととも、古典的テスト理論に基づいて探索的因子分析を適用して、サッカー技能の因子構造に対するテストの構成概念妥当性 (因子妥当性) が明らかにされた。しかし、ゲームパフォーマンスへの妥当性が十分ではなかったことから、ゲームパフォーマンスからサッカー技能を測定する項目を開発することが研究課題となった。

妥当性のある新しいゲームパフォーマンス測定項目を開発するためには、先行研究を対象とする文献研究だけでは不十分である。毎シーズンに欧州チャンピオンズリーグが先導し、発展し続ける最先端のサッカープレーに関与する Technical and tactical skill (技術戦術技能) は、実践知や暗黙知としてサッカーの専門家が認識しており、これらの定性的知識を抽出することが必要である。このような課題を解決するために、定性的分析手法を適用した新しい尺度構成法により、サッカーのゲームパフォーマンスからタクティカルスキルやテクニカルスキルのサッカー技能を測定するための項目が開発された。

我々の研究グループでは、体力・運動能力テストと運動生活行動の質問紙テストのための多変量統計解析による尺度構成法と項目特性分析の技術を適用して、サッカーゲームパフォーマンスからサッカー技能を測定する項目を開発した。Suzuki and Nishijima^{29,30)} は、定性的分析法としてデルファイ法を伴う特性要因分析を適用して、サッカーの専門家が視認しているサッカーの攻撃と守備のゲームパフォーマンス項目を抽出した。サッカーゲームデータに尺度構成法を適用して、妥当性と信頼性のある攻撃技能と守備技能の測定項目を開発した。

Suzuki and Nishijima²⁹⁾ は、サッカーの攻撃局面のプレーに関与する「スペースをつくる」、「しかける」、「くずす」から構成される攻撃技能を測定するゲームパフォーマンス項目を開発した。

Suzuki and Nishijima (2004)³⁰⁾ は、戦術守備 (Delaying attack、Forcing play in one direction、Squeezing working space of attackers) と対人守備

(Defence to attacker with the ball、Defence to attackers without the ball、Defence to attacking space) から構成される守備技能を測定するゲームパフォーマンス項目を開発した。構造方程式モデリングを適用して、サッカーゲームにおける逐次的なプレーに関与するサッカーの攻撃技能と守備技能を測定するゲームパフォーマンス項目の妥当性を明らかにした。

Tenga et al.³¹⁾ は、文献研究から攻撃戦術 (Offensive tactics) 18 項目、守備戦術 (Defensive tactics) 3 項目、ポゼッション結果 (Possession outcome) 1 項目の全 22 項目を構成し、サッカーのチームパフォーマンス分析のための測定項目を開発した。

継続研究である Tenga et al.³²⁾³³⁾ では、プレー戦術 (Playing tactics) と攻撃結果 (Goal scoring、Score-box 侵入) の関係を明らかにした。相手チームの不安定な守備 (Imbalanced defense) に対するカウンターアタック (Counter attack)、ファイナルサード (Final third) で開始した攻撃、5 回以上のパスから構成されるロングポゼッション (Long possession) での攻撃、ペネトレーティブパス (Penetrative pass) での攻撃は、攻撃結果に関係することを明らかにした。

Lago-Ballesteros, and Lago-Peñas¹⁶⁾ と Lago-Ballesteros et al.¹⁷⁾ は、攻撃結果 (Score-box 侵入) へのプレー戦術 (Playing tactics) の効果を分析し、Score-box 侵入のためにダイレクトアタック (Direct attack) とカウンターアタック (Counter attack) は、エラボレーティブアタック (Elaborative attack) の 3 倍の効果がある攻撃戦術プレーであることを明らかにした。

我々の研究グループでは、定性的分析法と達成度評価のテスト理論を適用して、サッカーゲームにおけるプレー単位の達成度から守備と攻撃におけるサッカー技能を達成度評価する項目を開発した。Matsuoka et al.²²⁾²³⁾ では、定性的分析法としてデルファイ法を伴う特性要因分析を適用して、サッカーの専門家が視認している最新のサッカーゲームにおける守備と攻撃のプレー項目と達成基準を抽出した。サッカーゲームのトラッキングデータとボールイベントデータからプレー項目を構成し、妥当性と信頼性のあるサッカー技能の達成度評価項目を開発した。

Matsuoka et al.²²⁾ は、サッカーゲームのトラッキングデータとボールイベントデータからサッカーの守備プレー項目を構成し、サッカーゲームの守備技能の達成度評価項目を開発した。デルファイ法を伴う特性要因分析により抽出したサッカー守備プレーの定性的因果構造に対応する守備プレー項目を構成

した。構造方程式モデリングを適用して、Set defense、Control defense、Press defense から構成されるサッカー守備技能の因子構造から、トラッキングデータから構成された守備プレー項目の妥当性を明らかにした。決定木分析を適用して守備プレー項目の達成基準を明らかにし、トラッキングデータから構成した連続量の守備プレー項目を達成 2f 値データ (達成・未達成) に変換した。項目反応理論 (Item Response Theory: IRT) を適用して、守備プレー項目の項目特性とテスト特性を明らかにし、サッカーゲームの守備技能の達成度評価項目を開発した。

Matsuoka et al.²³⁾ は同様な手続きを適用して、サッカーゲームの守備への切替技能 (Defensive transition skill) を達成度評価する 16 項目を開発した。守備への切替プレー成否 (プレス成功の切替プレー、相手の遅攻に対する切替プレー、相手の速攻に対する切替プレー、守備への切替プレー) の達成度評価得点を比較し、守備への切替技能得点の妥当性を明らかにした。

Jo et al.¹³⁾ は、サッカーゲームのボールイベントデータからサッカー攻撃プレーを測定する 45 項目構成し、決定木分析を適用して、攻撃プレー項目の達成基準を明らかにした。攻撃プレーの達成 2 値データ (達成・未達成) に、探索的因子分析と検証的因子分析を適用した。得られた因子構造から、ボールイベントデータから測定されるサッカー攻撃技能は、Forward moving、Pass、Connection movement、Sideward moving、Width and depth、Side attack、Final Area であることを明らかにした。

以上のように、従来のサッカーのゲーム分析手法から測定されるゲーム観察データに因子分析や構造方程式モデリングの多変量解析を適用した尺度構成手法から、サッカーゲームから計測されたトラッキングデータやゲームイベントデータのようなビッグデータに達成度評価のテスト理論である項目反応理論分析を適用した達成度評価項目の尺度構成へと発展している。これらの項目はサッカーゲームにおけるプレーの達成度を測定することから、プレー単位の達成度評価データから、データ主導型コーチングが進展することが期待される。

文献

- 1) Barreira D, Garganta J, Guimaraes P, Machado J C, and Anguera M T (2014): Ball recovery patterns as a performance indicator in elite soccer. Proc. Inst. Mech. Eng. P. J. Sport Eng. Technol., 228: 61-72.
- 2) Brooks J, Kerr M, and Guttag J (2016): Using

- machine learning to draw inferences from pass location data in soccer. *Stat. Anal. Data Min.*, 9: 338–349.
- 3) Carling C, Bloomfield J, Nelsen L, and Reilly T (2008): The role of motion analysis in elite soccer. *Sports Med.*, 38: 839–862.
 - 4) Decroos T, Bransen L, Van Haaren J, and Davis J (2019): Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1851–1861.
 - 5) Folgado H, Lemmink K A, Frencken W, and Sampaio J (2014): Length, width and entroid distance as measures of teams tactical performance in youth football. *Eur. J. Sport Sci.*, 14: S487-S492.
 - 6) Forcher L, Altmann S, Forcher L, Jekauc D, and Kempe M (2022): The use of player tracking data to analyze defensive play in professional soccer - A scoping review. *Int. J. Sports Sci. Coach*, 0: 1–25.
 - 7) Frencken W, Poel H D, Visscher C, and Lemmink K (2012): Variability of inter-team distances associated with match events in elite-standard soccer. *J. Sports Sci*, 30 (12): 1207–1213.
 - 8) Goes F R, Kempe M, Meerhoff L A, and Lemmink K A (2019): Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. *Big data*, 7: 57–70.
 - 9) Goes F R, Meerhoff L A, Bueno M J O, Rodrigues D M, Moura F A, Brink M S, Elferink-Gemser MT, Knobbe A J, Cunha S A, Torres R S, and Lemmink K A P M (2021): Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *Eur. J. Sport Sci.*, 21: 481–496.
 - 10) Herold M, Goes F, Nopp S, Bauer P, Thompson C, and Meyer, T (2019): Machine learning in men's professional football: Current applications and future directions for improving attacking play. *Int. J. Sports Sci. Coach.*, 14: 798–817.
 - 11) Hobbs J, Power P, Sha L, and Lucey P (2018): Quantifying the value of transitions in soccer via spatiotemporal trajectory clustering. *MIT Sloan Sports Analytics Conference*.
 - 12) Hughes M, and Lovell T (2019): Transition to attack in elite soccer. *J. Hum. Sport Exerc.*, 1: 236–253.
 - 13) Jo H, Matsuoka H, Ando K, and Nishijima T (2022a): Achievement standards and causal structure of offensive skill measurement items in Japan Professional Football League. *Football Science*, 19: 59–77.
 - 14) Jo H, Matsuoka H, Ando K, and Nishijima T (2022b): Construction of offensive play measurement items and shot prediction model applying machine learning in Japan Professional Football League. *Football Science*, 19: 1–21.
 - 15) Kempe M, Vogelbein M, Memmert D, and Nopp S (2014): Possession vs. direct play: evaluating tactical behavior in elite soccer. *Int. J. Sports Sci.*, 4 (6A): 35–41.
 - 16) Lago-Ballesteros J, and Lago-Peñas C (2010): Performance in team sports: Identifying the keys to success in soccer. *J. Hum. Kinet.*, 25: 85–91.
 - 17) Lago-Ballesteros J, Lago-Peñas C, and Rey E (2012): The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *J. Sports Sci.*, 30: 1455–1461.
 - 18) Link D, and Hoernig M (2017): Individual ball possession in soccer. *PloS one*, 12: e0179953.
 - 19) Low B, Coutinho D, Gonçalves B, Rein R, Memmert D, and Sampaio J (2020): A systematic review of collective tactical behaviours in football using positional data. *Sports Med.*, 50: 343–385.
 - 20) Lucey P, Bialkowski A, Monfort M, Carr P, and Matthews I (2015): Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In *Proc. 8th annual MIT Sloan sports analytics conference*, 1–9.
 - 21) Mackenzie R, and Cushion C (2013): Performance analysis in football: A critical review and implications for future research. *J. Sports Sci.*, 31: 639–676.
 - 22) Matsuoka H, Tahara Y, Ando K, and Nishijima T (2020): Development of criterion-referenced measurement items for soccer defensive tactical play from tracking data. *Football Science*, 17: 29–40.
 - 23) Matsuoka H, Tahara Y, Ando K, and Nishijima T (2021): Development of criterion-referenced measurement items of defensive transition in soccer games from tracking data. *Int. J. Sport Health Sci.*, 19: 87–97.
 - 24) Memmert D, and Rein R (2018): Match analysis, big data and tactics: current trends in elite soccer. *German Journal of Sports Medicine/Deutsche Zeitschrift für Sportmedizin*, 69: 65–71.

- 25) Pappalardo L, Cintia P, Rossi A, Massucco E, Ferragina P, Pedreschi D, and Giannotti F (2019): A public data set of spatio-temporal match events in soccer competitions, *Nature Scientific Data*, 6: 236.
- 26) Rein R, and Memmert D (2016): Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. Springer Plus, 5: 1-13.
- 27) Sampaio J, and Maças V (2012): Measuring tactical behaviour in football. *Int. J. Sports Med.*, 33: 395-401.
- 28) Sarmiento H, Marcelino R, Anguera M T, Campaniço J, Matos N, and Leitão J C (2014): Match analysis in football: a systematic review. *J. Sports Sci.*, 32: 1831-1843.
- 29) Suzuki K, and Nishijima T (2002): Causal structure of the attacking skill in soccer games. *Japan J. Phys. Educ. Hlth. Sport Sci.*, 47: 547-567. (in Japanese except Abstract)
- 30) Suzuki K, and Nishijima T (2004): Validity of a soccer defending skill scale (SDSS) using game performances. *Int. J. Sport Health Sci.*, 2: 34-49.
- 31) Tenga A, Kanstad D, Ronglan L T, and Bahr R (2009): Developing a New Method for Team Match Performance Analysis in Professional Soccer and Testing its Reliability. *Int. J. Perform. Anal. Sport.*, 9: 8-25.
- 32) Tenga A, Holme I, Ronglan L T, and Bahr R (2010a): Effect of playing tactics on goal scoring in Norwegian professional soccer. *J. Sports Sci.*, 28: 237-244.
- 33) Tenga A, Holme I, Ronglan L T, and Bahr R (2010b): Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *J. Sports Sci.*, 28: 245-255.
- 34) Wagenaar M, Okafor E, Frencken W, and Wiering M A (2017): Using Deep Convolutional Neural Networks to Predict Goal-scoring Opportunities in Soccer. *ICPRAM*, 448-455.