

トラッキングデータを用いたサッカーの 試合における戦況変化の抽出

神谷 啓太¹・中西 航²・泉 裕一朗¹

(受付 2016 年 12 月 31 日；改訂 2017 年 3 月 1 日；採択 4 月 14 日)

要 旨

サッカーの試合において「戦況」とは両チームの攻守のやり取りの中で徐々に変化していくものである。戦況変化を自動的に抽出できれば試合を有利に進めるための戦略が立てられるだけでなく、観戦者に対する情報提供などの面で有用である。本研究では、戦況変化を選手やボール位置の時系列的な振る舞いの変化と考え、統計的变化検知手法である ChangeFinder を使用することでサッカーの試合における戦況変化の抽出を試みる。ChangeFinder とは 2 段階の VAR モデルのオンライン学習を行うことで非定常かつ多ノイズな時系列データへの変化検知が可能な手法である。トラッキングデータ等からボール位置、前線位置、コンパクトネス、守備脆弱度、攻撃率の 5 種類の指標の入力変数を作成した。実験の結果、検出された変化点に対応するように、その直前で VAR モデルパラメータの時系列的振る舞いに大きな変動が存在することが確認された。そのパラメータ変動から想定される戦況変化の内容は実際のプレー内容と概ね合致していたとともに、想定される戦況変化が検出できたことが確認された。

キーワード：サッカー、戦況、変化点検知、時系列分析、ChangeFinder。

1. はじめに

数多くのスポーツで、センサやデータ計測員を駆使して多量のデータを取得し、統計的手法によってそれらを分析し、得られた示唆や知見をチーム戦略や選手評価に活用することが一般的となっている。サッカーにおいても、ボール支配率やパス・シュート本数などの従来のスタッツを始め、ここ数年では選手の走った経路を逐次記録したトラッキングデータなどが計測されている。サッカーでは両チーム合わせ 22 人の選手とボールが互いに影響し合いながら 45 分ハーフを連続的にプレーしており、取得されるデータにはそのような複雑な動きがすべて記録されている。サッカーの試合の大局を俯瞰的に把握するためには、この複雑なデータから選手や観戦者にとって解釈可能な形で示唆や知見を発掘する研究や分析が望まれる。

関連する研究の一つとして、Kijima et al. (2014) はサッカーの試合展開を支配する単純な法則性に言及している。一見複雑に見える選手とボールの位置変化はフラクタル構造を持つというもので、サッカーの試合はある普遍的な法則に従うことを示唆している。しかし一方で、観戦者や選手が感じる「戦況」は試合中に大きく変化することがある。例えば、攻勢であったのにいつの間にか守勢に転じている、停滞していた状況が一気に動き出す、などである。このよう

¹ 東京大学大学院 工学系研究科：〒113-8656 東京都文京区本郷 7-3-1

² 東京工業大学 環境・社会理工学院：〒152-8552 東京都目黒区大岡山 2-12-1

な戦況変化は上記の不変性の崩れとも考えられ、戦況変化を適切に把握することができれば、試合を有利に進めることに繋がる。しかし現状、実際に試合を観戦もしくはプレーすること以外に戦況変化を把握する方法が存在しない。戦況変化を自動的に抽出できれば試合を有利に進めるための戦略が立てられるだけでなく、観戦者に対する情報提供などの面で有用であり、その研究意義は大きい。

これまで、取得されたデータは各選手のプレー特徴やパフォーマンスの定量的な評価・分析に活用されるなど、主として個別的な分析に活用されてきた (Football LAB, 2017; Gudmundsson and Wolle, 2014; Sarmiento et al., 2014)。しかしながら、22 人の選手とボールとの相互作用を考慮した大局的な分析はいまだ少ない。数少ない例として、Moura et al. (2012) や高橋 他 (2015) は、選手の相対的な位置関係から算出される指標を用いて、攻撃時と守備時との選手配置の相違や、局面での優勢・劣勢の判定を試みている。

これに対し本研究では、攻守の切り替えのような明らかに観測される事実よりも大局的な視点に立ち、上述の「戦況」の変化を統計学的に扱うことを考える。そこで、戦況変化の自動抽出に向け、まずは戦況変化とは何かを検討する必要があるが、上述した例はいずれも選手やボールの時系列的な振る舞いの変化した結果であると解釈できる。そのため、選手やボールに関するデータの時系列的振る舞いをモニタリングし、その振る舞いの変化した点を検出することで戦況の変化を抽出することが可能となると考えられる。以上を踏まえると、統計的变化点検出 (山西, 2009) の枠組みを使用し、データの時系列的な振る舞いの変化点を検出することでサッカーの試合における戦況変化の抽出を行うことができると考えられる。

なお、サッカーを始めとする集団競技の戦況は、選手個々の意思決定に依る動きだけでなく、チーム全体としての動き、もしくは選手相互間やボールとの関係から発生するものである。したがって戦況が代表されるようなある一つの変数を見つけることは難しく、複数の変数および変数間の関係に着目して分析する必要がある。また、実用を考える上で、リアルタイムに戦況変化が抽出されることが望ましい。すると、多次元時系列データを対象に、確率モデルとして時系列モデルを仮定して、リアルタイムに変化点度合いのスコアを計算していく手法である ChangeFinder (Takeuchi and Yamanishi, 2006) が適用可能であると考えられる。

本研究の目的は、トラッキングデータを用いてサッカーの試合における戦況変化の抽出を行うことである。第 2 章で ChangeFinder について説明した後、第 3 章で分析に用いる変数を選定する。第 4 章で実データに ChangeFinder を適用し、検出された戦況変化に関する解釈を行う。最後に、本研究のまとめと今後の課題を第 5 章でまとめる。

2. ChangeFinder を用いた戦況変化の抽出手法

本章では ChangeFinder を用いた戦況変化の抽出手法について説明する。まず、ChangeFinder を適用するにあたり必要となる VAR モデル (Vector Autoregressive model; 多変量自己回帰モデル) について確認した後、ChangeFinder の基本原理および VAR モデルのオンライン学習方法について説明する。

2.1 VAR モデル

まず、初期値の平均が μ である d 次元時系列変数 $\{x_t : t = 1, 2, \dots\}$ を K 次の VAR 過程によってモデル化すると、以下のように定式化される。

$$(2.1) \quad x_t = \sum_{i=1}^K w_i(x_{t-i} - \mu) + \mu + \varepsilon$$

ここで、 $\omega_i \in \mathbf{R}^{d \times d} (i = 1, \dots, K)$ は d 次元パラメータ行列であり、 ε は平均 0, 分散共分散行列 Σ のガウス分布 $\mathcal{N}(0, \Sigma)$ に従う確率変数である。

すると、上記の VAR モデルによって定式化された x_t に関する確率密度関数は以下のように表すことができる。

$$(2.2) \quad p(x_t|\theta) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x_t - \omega)^T \Sigma^{-1} (x_t - \omega) \right\}$$

なお、VAR モデルのパラメータをまとめて $\theta = \{\omega_1, \dots, \omega_K, \mu, \Sigma\}$ と表記し、 $\omega = \sum_{i=1}^K \omega_i (x_{t-i} - \mu) + \mu$ である。また、 T は転置を表す。

2.2 ChangeFinder

ChangeFinder では観測データに対し 2 段階の VAR モデルのオンライン学習を行う。1 段階目の VAR モデルの学習において観測値に対する外れ値スコアを計算し、その後、平滑化した外れ値スコアを入力とした 2 段階目の VAR モデルを学習することによって、変化点スコアの計算を行う機構となっている。すなわち、外れ値スコアが一定期間連続して高水準である場合に、観測データにおける変化を検出する手法と換言できる。本研究においては、選手やボールの時系列的な振る舞いを VAR モデルで表現し、その振る舞いが変化した点を検出することが目的である。選手やボールの振る舞いが一定期間連続してこれまでの振る舞いと異なっている場合に戦況が変化したと考えられるため、ChangeFinder の枠組みにおいて外れ値ではなく変化点を検知することが戦況変化を抽出することに相当する。

いま、時刻 $t-1$ までの観測値 x_1, \dots, x_{t-1} が得られているとする。すると、後述する SDAR (Sequentially Discounting AR model learning) アルゴリズムと呼ばれるオンライン忘却型学習アルゴリズムを用いることで、時刻 $t-1$ における確率密度関数 $p_{t-1}(x)$ が推定される。ひとたび観測値 x_t が観測されると、時刻 t での外れ値スコアが対数損失によって以下のように算出される。

$$(2.3) \quad \text{Score}(x_t) = -\log(p_{t-1}(x_t))$$

次に、上記の方法によって算出した外れ値スコアに対して、以下の T 次移動平均を計算する。

$$(2.4) \quad y_t = \frac{1}{T} \sum_{i=t-T+1}^t \text{Score}(x_i)$$

新たに得られた時系列データ $\{y_t : t = 1, 2, \dots\}$ を VAR モデルで 2 段階目のモデル化を行い、再び SDAR アルゴリズムを用いて学習を行う。 y_t が得られた際に学習された確率密度関数を q_t とすると、 y_t の対数損失 $-\log(q_{t-1}(y_t))$ も 1 段階目と同様に算出される。

最後に、上記の対数損失に対し T' 次移動平均を計算した結果を時刻 t における変化点スコア $\text{Score}(t)$ とする。

$$(2.5) \quad \text{Score}(t) = \frac{1}{T'} \sum_{i=t-T'+1}^t \{-\log(q_{i-1}(y_i))\}$$

この変化点スコア $\text{Score}(t)$ が大きいほど時刻 t における状態変化の度合いが大きいことを意味する。

2.3 SDAR アルゴリズム

SDAR アルゴリズムでは、観測値 x_t が観測される度に VAR モデルのパラメータである

$\theta = \{\omega_1, \dots, \omega_K, \mu, \Sigma\}$ を学習する. この際, 忘却効果を取り入れることによって過去の観測で得られた情報の影響を徐々に低減していく. これにより, もともと定常過程の取り扱いを旨としている VAR モデルを, 形式的に非定常なモデルの学習へ拡張したものとなっている.

まず, 各種パラメータおよび統計量の初期値 $\hat{\mu}, \hat{\Sigma}, C_i (i = 1, \dots, K)$ を定める. ここで, $\{C_i : i = 1, \dots, K\}$ は自己共分散関数である.

x_t を観測する度に, 以下の更新式を計算する.

$$(2.6) \quad \hat{\mu} \leftarrow (1-r)\hat{\mu} + rx_t$$

$$(2.7) \quad C_j \leftarrow (1-r)C_j + r(x_t - \hat{\mu})(x_t - \hat{\mu})^T$$

上式における $r (0 < r < 1)$ が忘却パラメータであり, 新たなデータから計算された統計量と過去のデータから計算されている統計量の更新比を制御する. r が大きいほど忘却の度合いが大きいこととなる.

次に, 以下の自己共分散関数とパラメータ行列に関する YuleWalker 方程式を解く (北川, 2005).

$$(2.8) \quad \sum_{i=1}^K \omega_i C_{j-i} = C_j \quad (j = 1, \dots, K)$$

最後に, 上記の解を $\hat{\omega}_1, \dots, \hat{\omega}_K$ とおき, 以下を計算する.

$$(2.9) \quad \hat{x}_t \leftarrow \sum_{i=1}^K \hat{\omega}_i (x_{t-i} - \hat{\mu}) + \hat{\mu}$$

$$(2.10) \quad \hat{\Sigma} \leftarrow (1-r)\hat{\Sigma} + r(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T$$

観測値 x_t が観測される度に, 式(2.6)~(2.10)を繰り返す.

3. 入力変数の検討

3.1 使用したデータ

本研究で使用したデータは, 2015 明治安田生命 J1 リーグ 1st ステージ第 2 節鹿島アントラーズ対湘南ベルマーレ戦および第 17 節松本山雅 FC 対湘南ベルマーレ戦の計 2 試合に関して, 1/30 秒毎にパスやタックルなどボール周辺で発生したイベントおよびその発生時刻と位置を取得したボールタッチデータと, 1/25 秒毎に選手及び審判のピッチ上での位置を取得したトラッキングデータの 2 種類である. なお, これらのデータはデータスタジアム株式会社から提供を受けたものである.

3.2 入力変数の選定

VAR モデルの入力変数として用いる指標について検討を行う. まず, 戦況を表すのに十分な入力変数を用意できるよう, 試合全体の流れを表現すると思われる指標や, 各チームの攻勢・守勢を表す指標を複数作成した. 具体的には, 以下のような検討により (a) ボールの動き, (b) 選手の動き, (c) ボールポゼッションの変化, (d) 選手位置の変化というカテゴリを設け, 変数候補を選定した.

まず, 1 章で述べたとおり, 戦況は直接的にはボールや選手位置の時系列的振る舞いで表現されると考えられる. そこで, ボールの動きを表す変数として (a-1)「ボール位置」を候補とする. また, 選手の動きを表す変数として, (b-1)「各選手位置」(22 人)と, その集計量である (b-2)「各チーム選手位置の重心」, (b-3)「両チーム選手位置の重心」および既往研究 (Kijima et

al., 2014)で提案された(b-4)「前線位置」という指標を候補とする。

次に、観戦者は、両チームが攻撃あるいは守備を行っている度合いやそのプレー内容によって戦況を感じ取っていると考えられる。そこでの戦況の変化とは、例えば1章で述べたような、攻勢だった試合展開が守勢に変化したというものである。そこで、ボールポジションの変化を表す変数として(c)「攻撃率(詳細後述)」を候補とする。

さらに、サッカーの戦術についての書籍(松田 他, 2015; 西部, 2013; 西部・北, 2011)およびプレーヤーを含む複数の専門家へのヒアリングを通して、一般に「守備はコンパクトに、攻撃はワイドに」行うことが良いとされていることに着目した。「コンパクト」とはピッチ上で同チームの選手が近接していることで、特にピッチ前後方向の位置関係を指す。「ワイド」とはピッチ上で同チームの選手が散らばっていることで、特にピッチ左右方向の位置関係を指す。このことから、選手位置の変化を表す変数として、両チームそれぞれの(d-1)「オフサイドラインの座標」(d-2)「最前線の選手の座標」を候補とする。また、各チーム最前線の選手から最後方の選手までの距離を表す(d-3)「コンパクトネス(詳細後述)」, ピッチ左右方向の選手の幅を表す(d-4)「ワイドネス」も候補とする。さらに、コンパクトとワイドという相反する目標に対する攻守のせめぎ合いに注目した(d-5)「守備脆弱度(詳細後述)」を候補とする。

なお、戦況を生起するプレーは、実際には選手個人により行われているという観点からは、各選手のパスやシュートなどの「アクション」やボールを保持していないときの攻撃の一環としてのダッシュ(「スプリント」)も変数候補となり得る。ただし、今回は以下の理由から、これらの変数は候補としていない。まず、同一のアクションであっても、各時点で他に取り得たアクションの選択肢によってその意味合いは異なるため、選択肢集合の情報を有さずにアクションを変数とすると、結果の解釈上困難が生じる。また、スプリントを抽出するためには、目の前の相手に対する守備のような必要に迫られた走行との区別が必要であるが、各時刻の位置データのみからこれを区別することは容易ではないためである。

続いて、以上の変数候補から、実際に利用する変数の選定を行う。まず、VARのパラメータの推定可能性から、なるべく少ない変数でVARを構築することが望ましい。そこで、多大な次元増加をもたらす選手個人を示す変数は採用せず、それを代表する集計的な変数を用いることを前提とする。すなわち、(a)「ボール位置」、(b)選手の動きを代表する変数としての「各チーム選手位置の重心」「両チーム選手位置の重心」「前線位置」、(c)攻撃率、(d)選手配置を代表する変数としての「オフサイドラインの座標」「最前線の選手の座標」「コンパクトネス」「ワイドネス」「守備脆弱度」が変数候補として残る。

さらに、複数の変数を統合することで別の変数と同様の意味を表す場合や、同じカテゴリの変数間に相関がある場合が考えられる。次元削減の観点および共線性のない必要最低限な変数組という観点から、「オフサイドラインの座標」と「最前線の選手の座標」は同様の意味を表す「コンパクトネス」で代表させる。また、「ワイドネス」と「守備脆弱度」はどちらも選手の散らばり度合いを表す指標であり、実際にも相関を持っていることが確認されたため、攻守のせめぎ合いの状況をより明示的に表すであろう「守備脆弱度」で代表させる。同様に相関を持つ「各チーム選手位置の重心」、「両チーム選手位置の重心」および「前線位置」は、外れ値に強いと考えられる「前線位置」で代表させる。

以上の検討により、最終的に、ボール位置、前線位置、コンパクトネス、守備脆弱度、攻撃率の5種類の指標を分析に用いる変数として選定した。

3.3 入力変数の説明

以下、それらの指標について詳細な説明を行う。特に、前線位置、コンパクトネス、守備脆弱度については図1に概念図を示している。なお、提供を受けたデータは1/25秒や1/30秒間

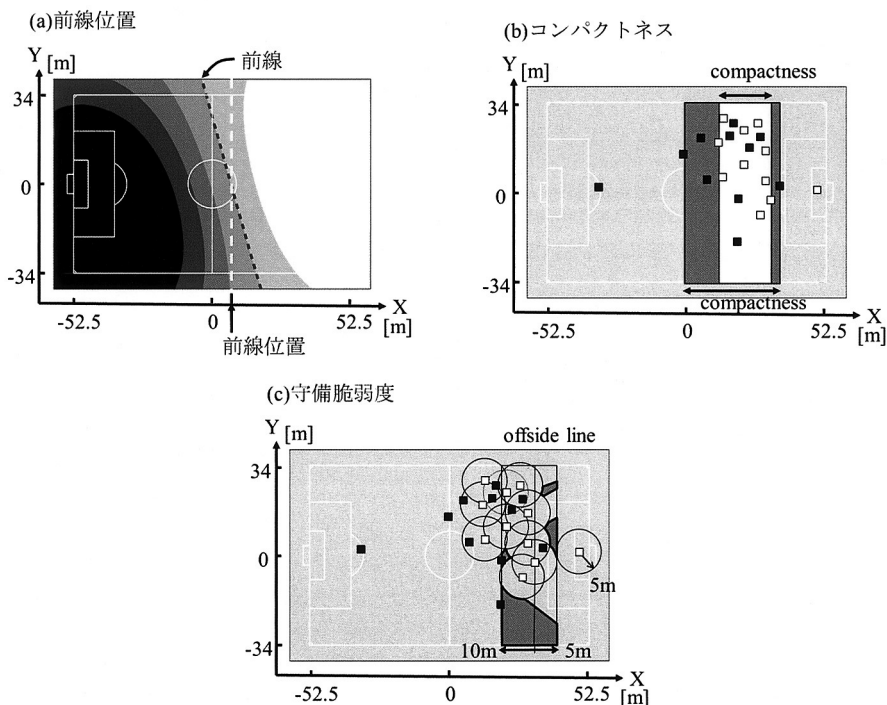


図 1. 入力変数の概念図。(a) 前線位置. 両チームの選手がなす支配領域が均衡する前線の X 座標に関する平均値. 図中の白色破線に対応。(b) コンパクトネス. 一番前方の選手と後方 2 番目の選手の X 座標上の距離. 図中の灰色と白色の四角形の幅にそれぞれ対応。(c) 守備脆弱度. 自軍のオフサイドラインより前方 10m, 後方 5m の長方形のうち, 最寄りの味方選手から 5m 以上離れており, 最近傍選手が相手選手であるような地点の合計面積の割合. 図中の灰色で囲まれている図形の合計面積が占める割合に相当.

隔であったが, VAR モデルへの適用を踏まえ, 分析に使うデータはすべて 1 秒間隔となるよう加工した. また, ピッチ中央を原点とし, コートの長辺方向を X 方向, 短辺方向を Y 方向と定義する.

まず, 試合全体の流れを表現する指標としてボール位置を選定した. 攻守の切り替わりによるボールの上下動のみを考慮するとし, X 座標のみを採用した. 時刻 t におけるボール位置を $ball(t)$ とする.

前線位置とは, 両チームの選手がなす支配領域が均衡する X 座標として定義した. この変数は全選手の動きを代表するため, 各チームの攻勢と守勢に合わせ値が増減すると予測される. 既往研究 (Kijima et al., 2014) に従い, 以下のとおり作成した.

- i) チームで正負の異なるガウスクERNEL (標準偏差 3m) を各選手が時刻 t に位置する座標に設定する.
- ii) 全てのKERNELを足し合わせ, ピッチ上で値が 0 となる線分を前線とする.
- iii) 前線の X 座標に対する平均値を算出し, これを時刻 t における前線位置とする.
- iv) i)~iii) を全時刻について算出し, 時刻 t における前線位置を $frontLine(t)$ とする.

コンパクトネスに関しては、ホーム・アウェイチーム毎に「一番前方に存在する選手の X 座標」と「後方より 2 番目に存在する選手の X 座標」の距離として定義し、時刻 t において $compact(HOME)(t)$ および $compact(AWAY)(t)$ とする。各チームの選手がその時刻において、どれだけピッチ上に広く存在しているかを示す指標である。攻撃時はピッチを広く使うためコンパクトネスは大きくなり、守備時は相手のプレーエリアを狭めるべくコンパクトネスは小さくなると予想される。

チームの守備力の脆弱性を表す指標として、自陣の最終ライン付近において相手選手が侵入している程度を守備脆弱度と定義し、時刻 t において $defense(HOME)(t)$ および $defense(AWAY)(t)$ とする。具体的には図 1(c) に示すように、「自軍のオフサイドラインより前方 10m, 後方 5m の長方形のうち、最寄りの味方選手から 5m 以上離れており、最近傍選手が相手選手であるような地点の合計面積の割合」である。攻撃時にボールを失って守備へ転換した直後には守備脆弱度は上昇し、守備陣営が整うと守備脆弱度は減少すると予想される。一方、カウンターや人数をかけた攻撃時には守備脆弱度は上昇することが予想される。

最後に、ホームチームの直近 10 分間における攻撃頻度を表す指標として攻撃率 $Attack(t)$ を定義する。ボールタッチデータ中に離散的に取得されている各チームの攻撃アクションを内挿し、1 秒毎の連続的な値として算出した。

$$(3.1) \quad Attack(t) = \frac{1}{6000} \int_{t-6000}^t f(t') dt'$$

$$(3.2) \quad f(t) = \begin{cases} 1 & \text{if } t \geq 0 \text{ and } A(t) = HOME \\ 0 & \text{if } t \geq 0 \text{ and } A(t) = AWAY \\ 0.5 & \text{if } t < 0 \end{cases}$$

ただし、 $A(t)$ は $t \geq 0$ において時刻 t 以前の最新の攻撃を行ったチーム (HOME または AWAY) を示す。

4. 戦況変化の抽出実験

4.1 適用条件

ChangeFinder の入力変数として、ボール位置、前線位置、コンパクトネス(ホーム・アウェイチームごと)、守備脆弱度(ホーム・アウェイチームごと)、攻撃率の 5 種類・計 7 つのデータを 7 次元変数として設定した。また、VAR モデルの次数 K は、オフラインでの VAR 推定を行った場合の最小 AIC 値を参考に、 $K = 5$ と決定した。忘却率 r や平滑化パラメータ T, T' についてはそれぞれ $r = 0.01$, $T = 50$, $T' = 5$ と設定した。鹿島・湘南戦および松本・湘南戦の前後半毎に ChangeFinder を適用し、戦況変化の抽出を試みる。

なお、SDAR アルゴリズムで用いる VAR モデルの各種パラメータおよび統計量の初期値 $\hat{\mu}, \hat{\Sigma}, C_i (i = 1, \dots, K)$ は、前後半それぞれの開始 60 秒間で得られたデータから計算することとし、この間の変化点スコアは算出しない。

4.2 適用結果および結果の考察

松本・湘南戦後半を例として、図 2 の上段に入力時系列データ x_t を、下段に ChangeFinder による変化点スコアの出力値 $Score(t)$ を示す。図 2 の上段において、1 行目にボール位置と前線位置、2 行目に湘南ベルマーレと松本山雅 FC のコンパクトネス、3 行目に湘南ベルマーレと松本山雅 FC の守備脆弱度がそれぞれ灰色の太線および黒色の細線で、そして 4 行目に松本山雅 FC の攻撃率が示されている。ボール位置と前線位置に関しては、正方向に値が上昇するに

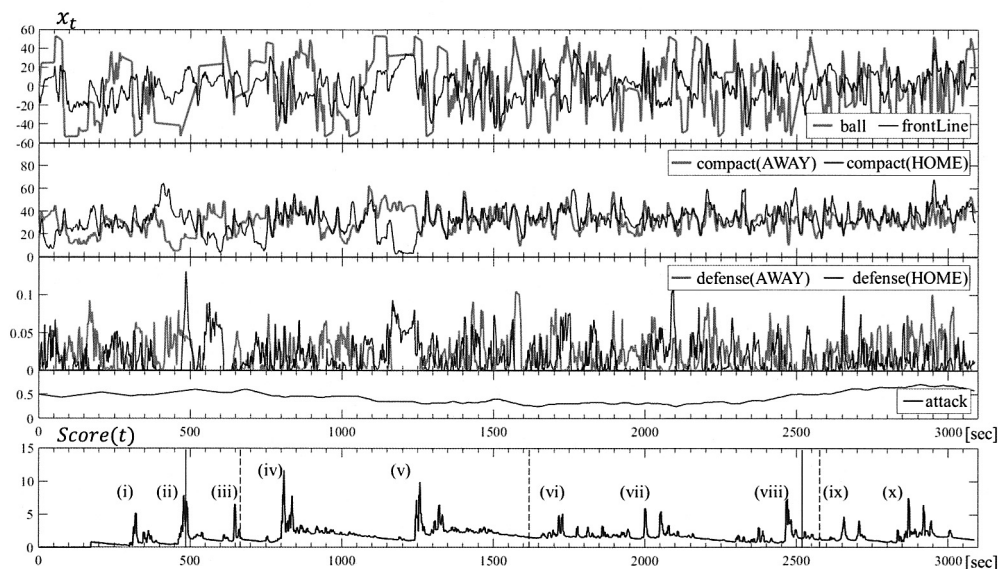


図 2. 松本・湘南戦後半における入力データおよび変化点スコア出力結果. 上段 1 行目にボール位置 (ball) と前線位置 (frontLine), 2 行目に湘南ベルマーレと松本山雅 FC のコンパクトネス (compact), 3 行目に湘南ベルマーレと松本山雅 FC の守備脆弱度 (defense) をそれぞれ灰色の太線および黒色の細線で, 4 行目に松本山雅 FC の攻撃率 (attack) を示す. AWAY および HOME はそれぞれ湘南ベルマーレと松本山雅 FC に対応する. 下段は ChangeFinder による変化点スコアの出力値 $Score(t)$ である. また, 下段グラフ中の実線および破線は, 松本および湘南による得点が生じた時刻である.

つれ, 湘南ベルマーレの攻撃方向に進出していることを意味する. 下段は各時刻について算出された変化点スコア $Score(t)$ であり, 値が大きいほど戦況変化の度合いが大きいことを表している.

出力された変化点スコアを見ると, 図中に (i)~(x) で示す合計 10 箇所では比較的大きな戦況変化が発生したと推定されたことが分かる. 紙面には掲載していないが鹿島・湘南戦前後半および松本・湘南戦前半の結果と比較すると, 松本・湘南戦後半ではより高頻度で変化点が検出されていた. 松本・湘南戦後半では, 53 分・87 分に松本が 2 点, 56 分・72 分・88 分に湘南が 3 点を獲得している. 片方のチームの得点の直後にもう片方のチームが得点するなど, 戦況の移り変わりが比較的大きかった試合であったことが, 本結果において比較的高頻度に変化点が検出された理由として考えられる. なお, 図 2 下段には松本による得点に対応する時刻を実線で, 湘南によるものを破線で示した. ゴールおよびその周辺の時刻にて変化点スコアが上昇していることが分かる.

次に, 検出された変化点が具体的にどういった戦況の変化を表しているのか, 入力データの振る舞いと実際のプレーを比較することで結果の解釈および検証を行う. ただし, 図 2 から分かるように入力データは複雑な動きをしているため, 入力データから変化点検出の原因を直接分析するのは困難である. そこで, パラメータ μ の各時刻における推定値 $\hat{\mu}$ に着目する. このパラメータは VAR モデルの式 (2.1) より, 自己回帰分を除いた平均的な値と解釈できる. この推定値 $\hat{\mu}$ の振る舞いに着目し, 検出された変化点との関係を分析することによって, どういった戦況の変化を表しているのか解釈を進めることが可能になると考える.

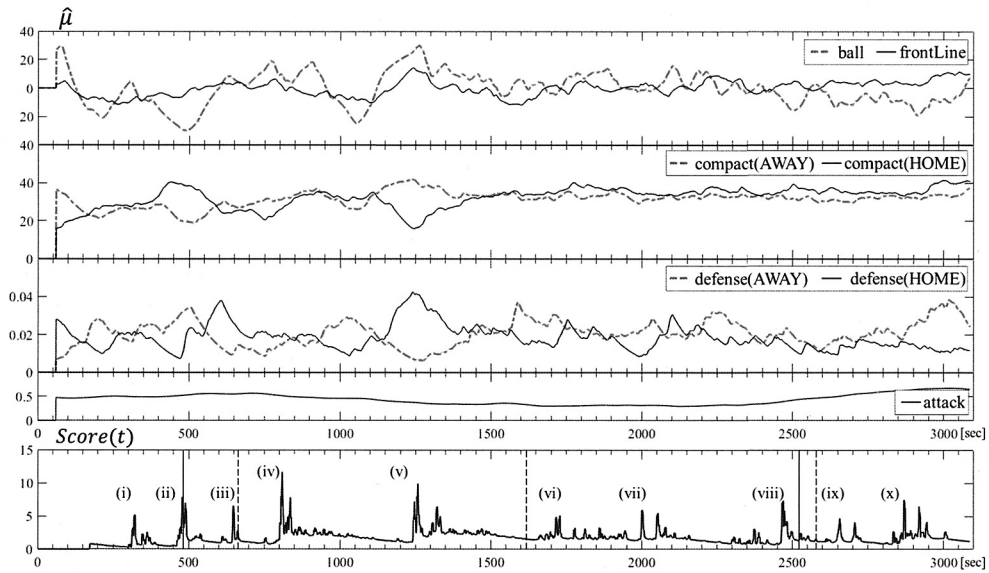


図 3. 松本・湘南戦後半におけるパラメータ μ の推定値 $\hat{\mu}$ の推移および変化点スコア出力結果。図中のグラフと変数の対応は図 2 に同じ。ただし、図 2 における灰色の太線が、本図では灰色の破線へ置き換わっている。また、試合開始 60 秒間のパラメータは推定されていない。

そこで、図 3 に松本・湘南戦後半における推定値 $\hat{\mu}$ の推移と変化点の検出結果を示す。図中のグラフと変数の対応は図 2 と同様である。ただし視認性の向上のため、図 2 における灰色の太線が、図 3 では灰色の破線へと置き換わっている。図 2 で複雑な振る舞いをしていた入力値が、図 3 の推定値 $\hat{\mu}$ では比較的滑らかな振る舞いをしていることが分かる。また、推定値の時系列的振る舞いに変動があると、それに対応するように直後 1 分後あたりで変化点スコアが上昇していることが分かる。なお、ChangeFinder では第 1 段階および第 2 段階で移動平均処理を行っているため、観測データの入力から変化点の検出までに遅延が生じることに注意が必要である。

表 1 に、各戦況変化点 (i) ~ (x) について代表的な $\hat{\mu}$ の変動の様子およびそれから想定される戦況変化の内容、ならびに実際にボールタッチデータから確認されたプレーの内容をまとめた。例えば戦況変化点 (i) の直前では、ボールが湘南攻撃方向に移動し、松本の守備脆弱度が上昇するという変動がパラメータ推定値 $\hat{\mu}$ の中で確認された。この結果より、湘南が攻勢へと転換したことが戦況変化の内容として想定される。実際にボールタッチデータより、湘南が複数のパスを繋ぎながら攻撃を展開している様子が確認できた。また、戦況変化点 (vii) では松本の守備脆弱度の低下およびその直後の上昇が確認されており、松本が陣営を立て直したうえで攻勢へと展開したことが想定される。実際のプレーを確認すると、1 点ビハインドである松本が飯尾 (MF) から石原 (MF) へ、喜山 (MF) から阿部 (FW) への選手交代を行っており、攻撃的な戦術へと変更した際の影響が検知されたものと考えられる。他の変化点についても、表中にあるように、変化点スコア上昇の原因となるようなパラメータ推定値 $\hat{\mu}$ の変動から想定される戦況変化の内容が、実際のプレー内容と概ね合致していることが確認できた。

以上の結果を踏まえると、本分析方法により、攻勢・守勢への転換やセットプレーによる連

表 1. 検出された変化点の解釈および実際のプレーとの関係.

#	代表的な $\hat{\mu}$ の変動	想定される戦況変化	実際に確認されたプレー
i	ボールが湘南攻撃方向に移動 松本の守備脆弱度が上昇	湘南の攻勢への転換	湘南がパスを繋ぎ攻撃を展開
ii	ボールが松本攻撃方向に移動 湘南のコンパクトネスが低下 松本のコンパクトネスが上昇 湘南の守備脆弱度が上昇 松本の守備脆弱度が低下	松本の攻勢への転換	松本が PK を獲得
iii	湘南のコンパクトネスが上げ止まり 松本のコンパクトネスが下げ止まり 松本の守備脆弱度が回復	松本の守勢の実現	湘南が FK を獲得し松本が守備体勢に
iv	ボールが松本攻撃方向に回復 松本のコンパクトネスが上昇	松本の攻勢への転換	松本が自陣で FK を獲得しピンチを脱する
v	松本のコンパクトネスが低下 松本の守備脆弱度が上昇 湘南の守備脆弱度が低下	湘南の攻勢への転換 松本の守勢の実現	湘南の CK/FK による連続的な攻撃
vi	湘南の守備脆弱度が回復	湘南の守備の実現	湘南得点後の松本キックオフ
vii	松本の守備脆弱度が低下および上昇	松本の攻勢への転換	松本の複数選手の交代による 攻撃的な戦術への変換か
viii	ボール位置が短時間に前後	試合の活性化	松本得点直後に湘南の得点
ix	ボールが松本攻撃方向に移動 湘南の守備脆弱度が上昇	松本の攻勢への転換	松本のシュートを含む試合展開
x	ボールが松本攻撃方向に移動 湘南の守備脆弱度が上昇	松本の攻勢への転換	同点弾を狙った松本の攻撃

統攻撃、連続得点による試合の活性化や選手交代による戦術変更など、想定される戦況変化は概ね検出できていると考えられる。ただし、実際の試合映像などを確認し、検出された戦況変化が妥当なものであるのか、また、未検出となった戦況変化等が存在しないかを専門的知見と共に検討を重ねたい。また、現在は検出された変化点と戦況変化の原因と考えられる実際のプレー内容との関係をパラメータ μ のみで見ているが、戦況のより高度な解釈に向け、 μ 以外のパラメータについても精査を行う必要がある。さらに、検出された戦況変化に対応するプレーにはセットプレーや選手交代などアウトオブプレーが多く含まれている。実際、セットプレーを機に戦況が大きく変わることは多々あるが、変化点(i)や(x)などインプレー中の戦況変化の抽出も重要であろう。インプレー中の戦況変化の抽出のためには、ChangeFinder に適用する時系列データからアウトオブプレー中のデータを排除する、一つのインプレーデータ毎にChangeFinder を適用するなど、分析方法を修正することが解決策として考えられる。

5. おわりに

本研究では、統計的変化点検出手法である ChangeFinder をトラッキングデータに適用することで、サッカーの試合における戦況変化の抽出を行った。ChangeFinder の入力変数の検討にあたっては、変数間の相関分析を通して共線性のない必要最低限な変数組となるよう配慮し、ボール位置、前線位置、コンパクトネス、守備脆弱度、攻撃率の5種類の指標を選定した。松本山雅 FC 対湘南ベルマーレ戦後半のトラッキングデータに ChangeFinder を適用した結果では、検出された合計約 10 か所の変化点に対応するように、その直前でパラメータ μ の推定値 $\hat{\mu}$ の時系列的振る舞いに変動が確認された。また、変化点スコア上昇の原因となるパラメータ推定値 $\hat{\mu}$ の変動から想定される戦況変化の内容と、実際のプレー内容がおおむね合致している

ことが確認できた。本研究は、選手交代のような戦術変更が実戦況にどのような効果を与えたかを測定・分析するツールとしての活用可能性も考えられる。

今度の課題として、実際の試合動画を専門的知見をもとに観察し、検出された戦況変化が妥当かつ十分なものであるか検証を進めたい。また、より高度な戦術分析の実現に向け、VARモデルの μ 以外のパラメータについても精査し、変化点が検出された原因について分析を重ねる必要がある。検出された変化点の前後でどのような戦況であったのか記述を行うことも今後の課題として挙げられる。そして、データの背後に潜む戦況に関する潜在構造を推定し、戦況を変化させる因子の検出が可能となれば、より高度な戦術分析の実現へ向け前進するであろう。

謝 辞

本研究で使用したデータはデータスタジアム株式会社から提供を受けたものである。また、本研究の貸与データは情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援を受けている。なお、本分析を進める中で、矢野慎一氏(東京大学・当時)にサッカーにおける専門的知見を提供いただいた。

参 考 文 献

- Football LAB (2017). Football LAB (フットボールラボ)とは, <http://www.football-lab.jp/pages/about/>, 2017年2月28日閲覧.
- Gudmundsson, J. and Wolle, T. (2014). Football analysis using spatio-temporal tools, *Computers, Environment and Urban Systems*, **47**, 16–27.
- Kijima, A., Yokoyama, K., Shima, H. and Yamamoto, Y. (2014). Emergence of self-similarity in football dynamics, *The European Physical Journal B*, **87**(2), 1–6.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- 松田浩, 鈴木康浩 (2015). 『サッカー守備戦術の教科書 超ゾーンディフェンス論』, カンゼン, 東京.
- Moura, F. A., Martins, L. E. B., Anido, R. D. O., De Barros, R. M. L. and Cunha, S. A. (2012). Quantitative analysis of Brazilian football players' organisation on the pitch, *Sports Biomechanics*, **11**(1), 85–96.
- 西部謙司 (2013). 『サッカーで大事なことは、すべてゲームの中にある：クローズアップ 世界トップクラスの技術とアイデア』, 出版芸術社, 東京.
- 西部謙司, 北健一郎 (2011). 『サッカー戦術とは何か? が誰でも簡単に分かるようになる本』, 毎日コミュニケーションズ, 東京.
- Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N. and Leitão, J. C. (2014). Match analysis in football: A systematic review, *Journal of Sports Sciences*, **32**(20), 1831–1843.
- 高橋翔, 長谷山美紀 (2015). サッカー映像におけるグループ戦術解析のための優先度推定に関する一考察, 映像情報メディア学会技術報告, **39**(49), 7–12.
- Takeuchi, J. and Yamanishi, K. (2006). A unifying framework for detecting outliers and change points from time series, *IEEE Transactions on Knowledge and Data Engineering*, **18**(4), 482–492.
- 山西健司 (2009). 『データマイニングによる異常検知』, 共立出版, 東京.

Tracking Data to Extract Changes in Football Game Situation

Keita Kamiya¹, Wataru Nakanishi² and Yuichiro Izumi¹

¹Department of Civil Engineering, The University of Tokyo

²Department of Civil and Environmental Engineering, Tokyo Institute of Technology

In football, a “game situation” gradually changes the interaction of teams’ attacks and defenses. The ability to automatically extract a change in a football game will facilitate the development of advanced strategies as well as provide richer information to the spectators. In this research, we regard the change in a football game situation as a change in time series behavior of players and a ball.

We attempt to extract the change using ChangeFinder, which is a statistical change detection method. ChangeFinder can detect changes in nonstationary and multi-noise time-series data via online learning of the two-step VAR model. Input variables are created from the tracked data to create five types of indicators: ball position, front line position, compactness, defense vulnerability degree, and attack rate. The experiments confirm that a large fluctuation in the time-series behavior of VAR model parameters occurs just prior to a detected change point. The contents of the change in a game situation assumed from the parameter variation roughly agree with the actual play contents, suggesting that the model can detect changes in football game situations.