

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（東京大学大学院工学系研究科）

中西 航（東京大学大学院工学系研究科・助教）

泉 裕一郎（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

TEL:03-5841-6118, FAX:03-5841-7453

E-MAIL:kamiya@trip.t.u-tokyo.ac.jp

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 ら [1] 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。

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、統計的変化

点検出 [2] 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[3] 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 $\{\mathbf{x}_t : t = 1, 2, \dots\}$ 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。

$$\mathbf{x}_t = \sum_{i=1}^K w_i(\mathbf{x}_{t-i} - \mu) + \mu + \varepsilon \quad (1)$$

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 \mathbf{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。

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

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、 $\omega = \sum_{i=1}^K \omega_i(\mathbf{x}_{t-i} - \mu) + \mu$ 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。

いま、時刻 $t-1$ までの観測値 $\mathbf{x}_1, \dots, \mathbf{x}_{t-1}$ 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 $p_{t-1}(\mathbf{x})$ が推定される。ひとたび観測値 \mathbf{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。

$$Score(\mathbf{x}_t) = -\log(p_{t-1}(\mathbf{x}_t)) \quad (3)$$

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。

$$y_t = \frac{1}{T} \sum_{i=t-T+1}^t Score(\mathbf{x}_i) \quad (4)$$

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 $Score(t)$ とする。

$$Score(t) = \frac{1}{T'} \sum_{i=t-T'+1}^t \{-\log(q_{i-1}(y_i))\} \quad (5)$$

この変化点スコア $Score(t)$ 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 \mathbf{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を徐々に低減していく。これにより、もともと定常過程のみを取り扱うことができる VAR モデルが、非定常なモデルの学習も形式的と可能となっている。

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。

\mathbf{x}_t を観測する度に、以下の更新式を計算する。

$$\hat{\mu} \leftarrow (1-r)\hat{\mu} + r\mathbf{x}_t \quad (6)$$

$$C_j \leftarrow (1-r)C_j + r(\mathbf{x}_t - \hat{\mu})(\mathbf{x}_{t-j} - \hat{\mu})^T \quad (7)$$

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 [4]。

$$\sum_{i=1}^K \omega_i C_{j-i} = C_j \quad (j = 1, \dots, K) \quad (8)$$

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。

$$\hat{\mathbf{x}}_t \leftarrow \sum_{i=1}^K \hat{\omega}_i(\mathbf{x}_{t-i} - \hat{\mu}) + \hat{\mu} \quad (9)$$

$$\hat{\Sigma} \leftarrow (1-r)\hat{\Sigma} + r(\mathbf{x}_t - \hat{\mathbf{x}}_t)(\mathbf{x}_t - \hat{\mathbf{x}}_t)^T \quad (10)$$

観測値 \mathbf{x}_t が観測される度に、式 (6)~(10) を繰り返す。

3. 入力変数の検討

3.1. 使用したデータの説明

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。

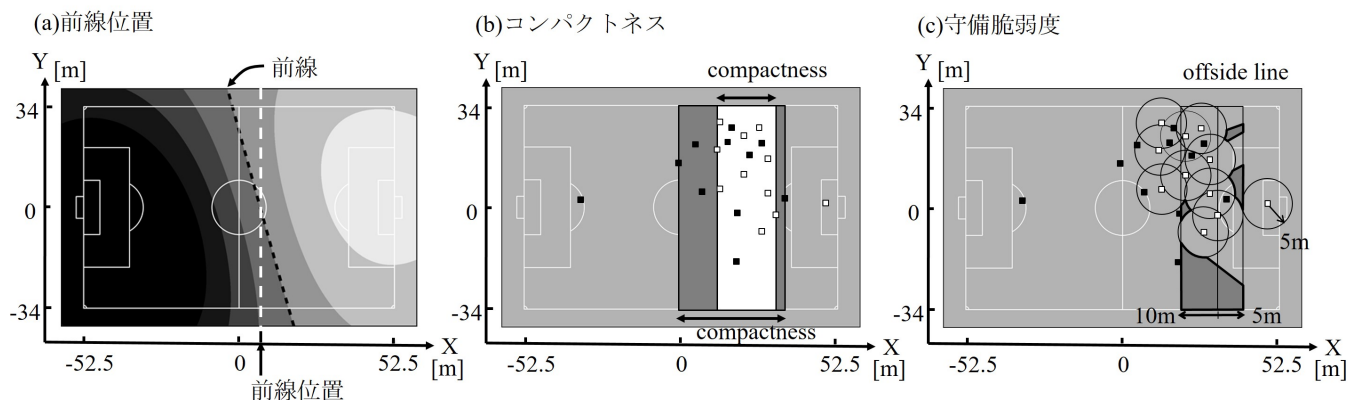


図 1: 入力変数の概念図. (a) 前線位置. 両チームの選手がなす支配領域が均衡する前線の X 座標に関する平均値. 図中の白色破線に対応. (b) コンパクトネス. 一番前方の選手と後方 2 番目の選手の X 座標上の距離. 図中の灰色と白色の四角形の幅にそれぞれ対応. (c) 守備脆弱度. 自軍のオフサイドラインより前方 10m、後方 5m の長方形のうち、最寄りの味方選手から 5m 以上離れており、最近傍選手が相手選手であるような地点の合計面積の割合. 図中の灰色で囲まれている図形の合計面積が占める割合に相当.

3.2. 入力変数の検討

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に、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。その後、変数間の相関分析を行い、共線性のない必要最低限な変数組が選定できるように配慮した。その結果、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。

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、コートの方を X 方向、短辺方向を Y 方向と定義する。

3.2.1 ボール位置

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。攻勢・守勢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。

3.2.2 前線位置

両チームの選手がなす支配領域が均衡する X 座標を前線位置と定義した。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 [1] に従い、以下のとおり作成した。

- チームで正負の異なるガウスカネル（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。
- 全てのカーネルを足し合わせ、ピッチ上で値が 0 となる線分を前線とする。

- 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。

- i) ~ iii) を全時刻について算出する。

3.2.3 コンパクトネス

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」の距離をコンパクトネスと定義する。各チームの選手がその時刻において、どれだけピッチ上で展開できているかを示す指標であり、攻撃を展開するとコンパクトネスは大きくなり、守備に回るとコンパクトネスは小さくなると予想される。

3.2.4 守備脆弱度

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。守勢への転換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。

3.2.5 攻撃率

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。

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・湘南戦の前後半のハーフ毎に 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、ハーフ開始 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 \mathbf{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および黒色の実線で、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・湘南戦後半では両チーム合わせ5つのゴールが生まれており、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。

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からも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 (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性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。

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。図2で複雑な振る舞いをしていった入力値が、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。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続攻撃、連続得点による試合の活性化など、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、統計的变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

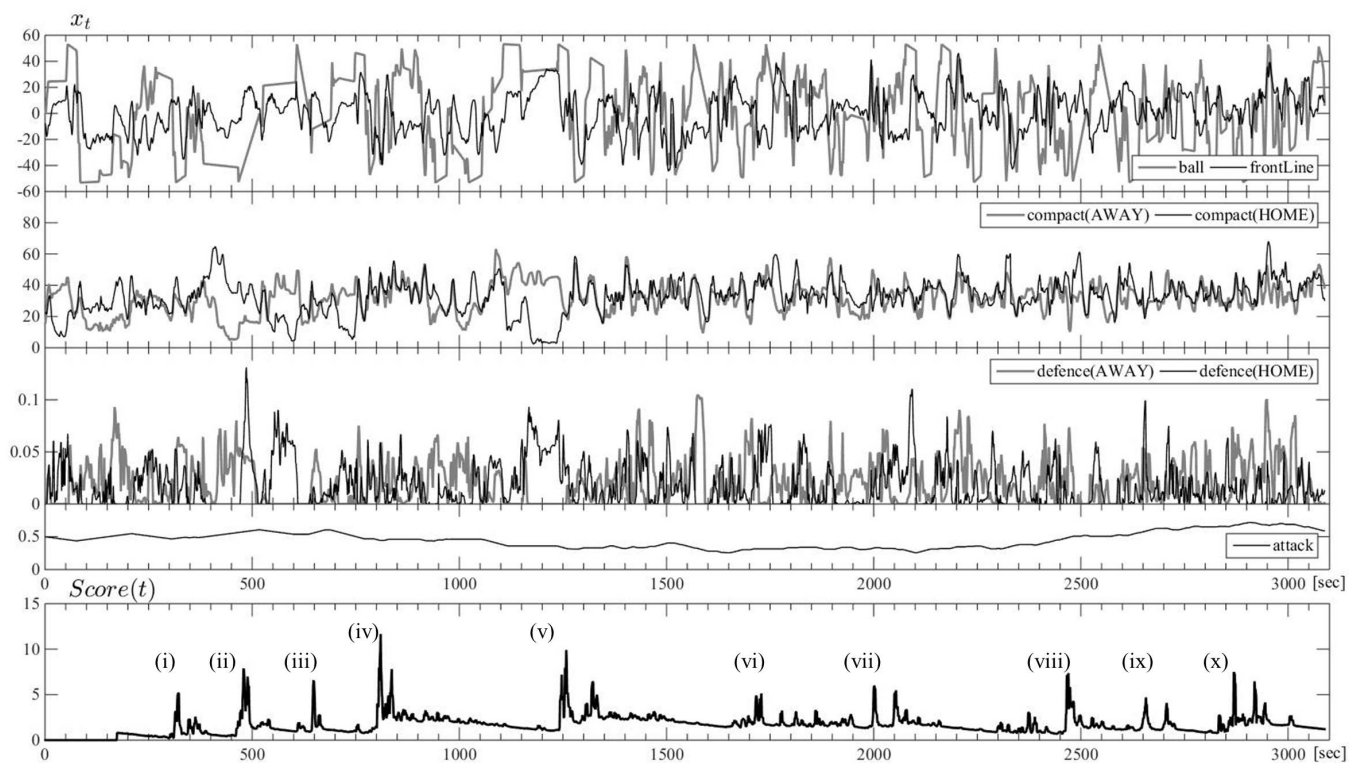


図 2: 松本・湘南戦後半における入力データおよび変化点スコア出力結果。 上段 1 行目にボール位置 (ball) と前線位置 (frontLine)、2 行目に湘南ベルマーレと松本山雅 FC のコンパクトネス (compact)、3 行目に湘南ベルマーレと松本山雅 FC の守備脆弱度 (defence) をそれぞれ灰色および黒色で、4 行目に松本山雅 FC の攻撃率 (attack) を示す。AWAY および HOME はそれぞれ湘南ベルマーレと松本山雅 FC に対応する。下段は ChangeFinder による変化点スコアの出力値 $Score(t)$ である。

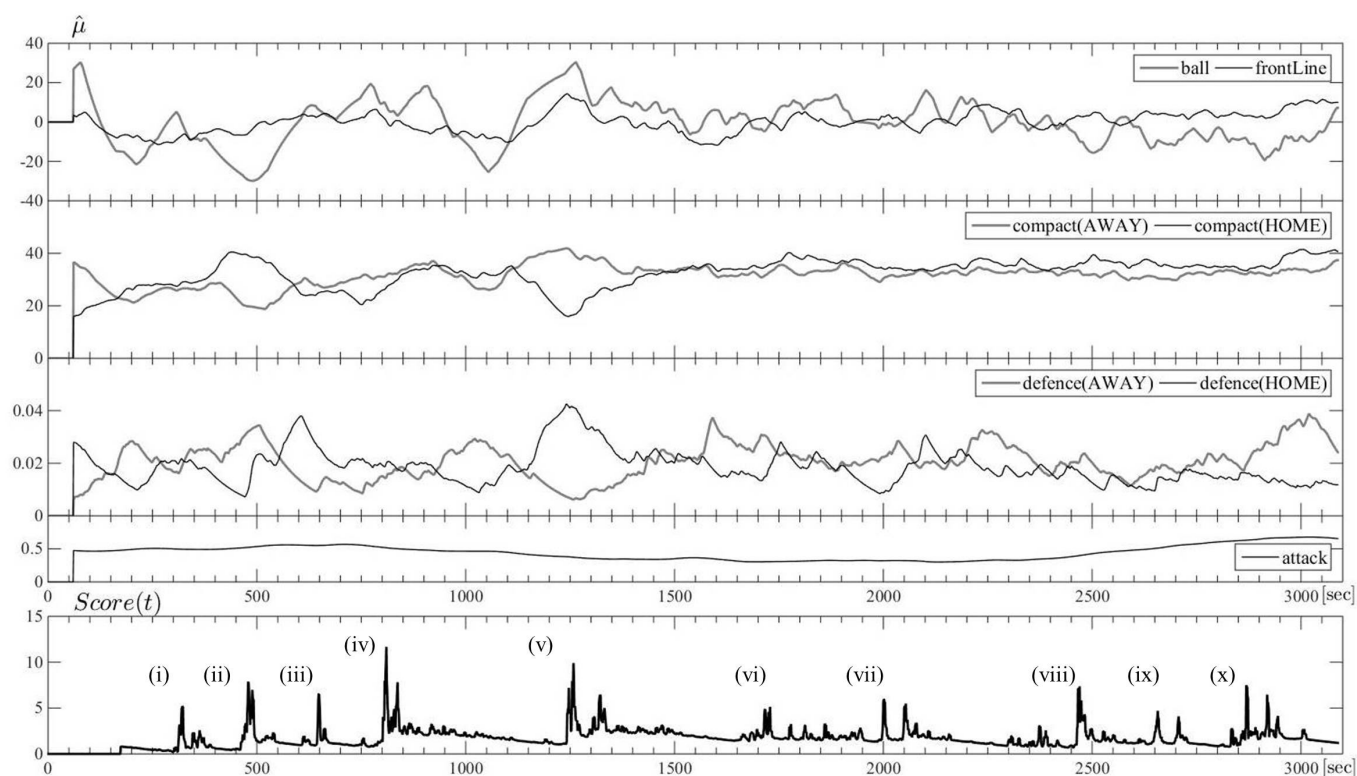


図 3: 松本・湘南戦後半におけるパラメータ μ の推定値 $\hat{\mu}$ の推移および変化点スコア出力結果。 図中のグラフと変数の対応は図 2 に同じ。なお、試合開始 60 秒間のパラメータは推定されていない。

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	松本の守勢の実現	複数選手の交代により松本守勢の立て直し
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

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、戦術変更が実戦に効果を与えたか測定・分析するためのツールとして有用であると考えられる。

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）にサッカーにおける専門的知見を提供いただいた。ここに、感謝の意を表する。

参考文献

- [1] A. Kijima, K. Yokoyama, H. Shima and Y. Yamamoto, “Emergence of self-similarity in football dynamics,” The European Physical Journal B, vol. 87, no. 2, pp. 1-6, 2014 .
- [2] 山西健司「データマイニングによる異常検知」共立出版、2009.
- [3] Takeuchi, J., and Yamanishi, K., “A Unifying Framework for Detecting Outliers and Change Points from Time Series,” IEEE Trans. on Knowledge and Data Engineering, 18(4), pp.482-492, 2006.
- [4] 北川源四郎「時系列解析入門」岩波書店、2005.