

チームスポーツにおける数的優位に至る過程の分析

Analyzing Processes of Creating Numerical Advantage in Team Sports

田中佑典^{*1} 戸田浩之^{*1} 湯口昌宏^{*1}
Yusuke Tanaka Hiroyuki Toda Masahiro Yuguchi

^{*1}NTT サービスエボリューション研究所
NTT Service Evolution Laboratories

Analyzing team tactics is important for the decision making of soccer coaches. With the recent advances in information technology, the mobility of ball and players can be tracked over time. This paper proposes a method for analyzing the processes that create numerical advantage in team sports. Our method contains the following three procedures: 1) extract the offensive patterns from ball trajectories, 2) split the offenses into two classes on the basis of creating numerical advantage, 3) use logistic regression with L_1 regularization. Experiments on real-world data collected from J-League show that our method can 1) extract the representative offensive patterns, and 2) find the factors of creating numerical advantage in soccer games.

1. はじめに

情報技術の発展により、サッカーやバスケットボールなど様々なスポーツにおいて、選手の詳細なプレイ履歴が大規模に蓄積されつつある。プレイ履歴に基づいて、選手の能力やチームの戦略を分析することは重要である。分析結果は、チームの戦略策定 [Lucey 13] や選手の評価 [Brooks 16] などに利用されている。

選手のプレイログ（シュートやパスなどの行動履歴）と選手の軌跡情報とに基づいて、チームの攻撃パターンの分析がなされている [Lucey 13, Wei 13]。選手間のパスの頻度とプレイがなされた位置によって攻撃パターンを定義し、トピックモデルなどを応用することにより、チームの頻出攻撃パターンを抽出する手法 [Wang 15] が提案されている。抽出された各攻撃パターンに対し、シュート成功率を計算することによってチームが得意とするパターンを分析することも可能である。しかし、従来の研究では、攻撃が成功するに至った過程や、その要因についての分析は十分に行われていない。得意（不得意）な攻撃パターンの抽出に留まらず、攻撃が“なぜ”成功（失敗）したのか、について分析することは、戦略策定や選手の評価を行う際に重要な指針となりうる。

本研究では、サッカーにおける選手のプレイ履歴に基づいて、攻撃が成功に至る過程や要因についての分析を行うことを目的とする。この目的を達成するためには、解決すべき三つの問題が存在する。一つ目は、試合中には複数の攻撃パターンが混在するため、これらをいかに分離するかという問題である。例えば、コート中央を突破するような攻撃パターンがあれば、サイドからのセンタリングや、コーナーキックからのヘディングシュートのような攻撃パターンなどが考えられる。二つ目は、攻撃の成否をどのように評価するかという問題である。最も単純には、シュート成功を攻撃の成功とみなすことが考えられる。しかし、シュートの成否は選手個人の能力に左右されることから適切でない場合がある。そこで、攻撃の成否を評価するための指標について検討することは重要である。三つ目は、攻撃の成功に至る過程をどのように分析するかという問題である。攻撃の成功に至る要因は、直前の選手の行動のみに依存するとは限らない。例えば、シュートから少し前の時

表 1: 選手のプレイログ

攻撃 ID	時刻	選手	行動	座標 (x,y)
1	01:13	選手 A	パス	(116.6, 162.6)
1	01:17	選手 B	ドリブル	(126.9, 132.6)
1	01:21	選手 C	シュート	(131.3, 122.1)
2	02:03	選手 B	パス	(562.6, 100.5)
2	02:09	選手 A	パス	(121.6, 321.1)
⋮	⋮	⋮	⋮	⋮

刻におけるボール非保持者の走り込みが守備選手を引き付け、シュートスペースが作り出される場合などが考えられる。

上記の三つの問題を踏まえて、攻撃が成功に至る過程や要因について分析するための手法を構築する。分析手順は、(a) ボールの軌跡に基づきクラスタリング手法を用いて攻撃パターンを抽出し、(b) シュートの瞬間に攻撃側が数的優位な状況にあるか否かを指標として攻撃を二クラスに分類し、(c) L_1 正則化付きロジスティック回帰を用いて数的優位に至る過程について分析を行う。ここで、数的優位な状況とは、ボール位置より前（ボールと相手ゴールの間）にいる攻撃側の選手人数が、守備側の選手人数よりも多いことを意味する。本研究では、攻撃の成功を、数的優位な状況を作り出すことであると仮定し、その状況に至る過程や要因についての分析を試みる。

評価実験では、サッカーの実データを分析した結果について報告する。まず、ボールの軌跡に基づきクラスタリング手法を用いて攻撃パターンを抽出した結果を示す。次に、数的優位な状況を作り出せたか否かという指標が、攻撃の成否を評価するために妥当であることを定量的に示す。最後に、数的優位な状況に至る過程や要因についての分析結果を示す。

2. データ

本節では、サッカーにおけるプレイ履歴について説明する。本研究で用いるプレイ履歴は、選手のプレイログ（シュート、パス、ドリブルの 3 種類の行動履歴）と選手のトラッキングデータ（位置情報の時系列データ）とから構成される。表 1 に選手のプレイログの例を示す。プレイログは、各チームの攻撃

ID 毎のボール保持者に関する履歴である。一方、トラッキングデータはボール非保持者についても各時刻における座標情報が記録されている。x 座標はゴール方向を表し、取りうる範囲は $-5250 < x < 5250$ である。y 座標はゴール方向と直角の方向を表し、取りうる範囲は $-3400 < y < 3400$ である。次節では、プレイログおよびトラッキングデータを入力として分析を行う手法について説明する。

3. 分析手法

3.1 概要

本節では、選手のプレイログおよびトラッキングデータに基づいて、攻撃が成功に至る過程や要因についての分析を行うための手法を構築する。分析の手順は、(a) ボールの軌跡に基づきクラスタリング手法を用いて攻撃パターンを抽出し、(b) シュートの瞬間に攻撃側が数的優位な状況にあるか否かを指標として攻撃を二クラスに分類し、(c) L_1 正則化付きロジスティック回帰を用いて数的優位に至る過程について分析を行う。以下では、(a) および (c) の各処理について説明する。

3.2 攻撃パターンの抽出

各攻撃機会における選手のプレイログのみを入力として、複数の代表的な攻撃パターンを抽出する。攻撃機会とは、攻撃側のチームがボールを奪ってからシュートに至るまでの期間のことである。文献 [Wei 13] に倣って、表 1 に示したプレイログに含まれるボール軌跡に基づいて攻撃パターンの抽出を行う。プレイログにおいて、各攻撃機会に含まれるログの数が異なるため、ボール軌跡をベクトルとして表現し、そのベクトルを入力としてトピックモデルや k -means 法などのクラスタリング手法によって攻撃パターン抽出を行う方法が提案されている [Wei 13, Wang 15]。ボール軌跡のベクトル化に最もよく用いられる方法は、コートをもとに G 個のグリッドに分割する方法である [Wei 13]。各グリッドをボールが通った回数を計算することで、 G 次元のベクトルとすることができる。しかし、この方法ではコートをもとに分割するため、実際のボール位置を適切に表現できない場合がある。例えば、サッカーにおいてセンタリングをあげる場所が特定のエリアに偏っている場合は、それを考慮することはできない。そこで本研究では、ボールの位置情報に Mean Shift 法 [Comaniciu 02] を適用することによって、ボールの密度が高い G' 個の代表点を抽出する。そして、軌跡に含まれる各ボール位置を、最も近い代表点の一つに割り当てる。各代表点をボールが通った回数を計算することで G' 次元のベクトルを得る。

本研究では、上記のような方法で作成したベクトルを入力として、 k -means 法を適用することによって k 個の攻撃パターンを抽出する。

3.3 数的優位に至る過程の分析

3.2 節において抽出された各攻撃パターンにおいて、攻撃チームが数的優位に至る過程の分析を行う方法について述べる。まず、攻守両方の選手のトラッキングデータを用いて、各攻撃機会においてシュートの瞬間に数的優位な状況になっているか否かを判定する。ここで、数的優位な状況とは、ボール位置より前（ボールと相手ゴールの間）にいる攻撃側の選手人数が、守備側の選手人数よりも多いことを意味する。これにより、各攻撃機会に対してクラスラベル $y = \{0, 1\}$ を割り振る。 $y = 1$ のとき数的優位、 $y = 0$ のとき数的優位でないことを定義する。

次に、各攻撃機会に対して特徴量 x を計算する。実験に用いた特徴量については 4.2 節で述べる。特徴量の数を D とした

とき、 D 次元特徴ベクトルを $x = (x_1, \dots, x_D)^T$ と表す。また、シュートに至るまでの過程の分析を行うため、シュートからさかのぼって 3 ステップ前までの特徴ベクトルを作成する。シュートの時刻を t_0 、1 ステップ前の行動（パスやドリブル）の時刻を t_1 としたとき、 $t_1 < t \leq t_0$ に含まれる選手やボールの軌跡情報を用いて特徴量を計算し、シュートに至る 1 ステップ前の特徴ベクトル x_1 とする。同様にして 2、3 ステップ前の特徴ベクトルをそれぞれ作成し、まとめて $X = (x_1, x_2, x_3)$ とする。 (X, y) を入出力のペアとして、ロジスティック回帰を用いて識別器を学習する。 $y = 1$ である事後確率は、

$$P(y = 1|X) = \frac{1}{1 + e^{-(w_1^T x_1 + w_2^T x_2 + w_3^T x_3 + b)}} \quad (1)$$

と表される。ここで、 w_1, w_2, w_3 は各時間ステップにおける特徴ベクトルに対する重みベクトル、 b はバイアスを表す。学習の際には、 L_1 正則化項を加えて重みベクトルおよびバイアスの推定を行う。 L_1 正則化は、識別に有用でない重みベクトルの要素を 0 に近づける働きがあり、特徴選択の役割を果たす [Tibshirani 96]。これは識別に有用な特徴を分析する際に効果的である。

以上のように、数的優位な状況を作り出せたか否かの二値識別器を構成し、両者を分離するために有用な特徴を学習する。学習された重みベクトルの値を分析することによって数的優位な状況に至る過程や要因について分析することができる。

4. 実験

4.1 概要

本節では、サッカーの試合における実データを用いて攻撃パターンの分析を行った結果について述べる。4.2 節では、実験に用いたデータの詳細、および、分析に用いた特徴量について述べる。4.3 節では、ボール軌跡に基づいて攻撃パターンを抽出した結果について述べる。さらに、各攻撃機会に対し、数的優位な状況を作り出せたか否かを表す二値ラベルを割り当て、数的優位な状況が攻撃の成否を評価する指標として妥当であることを示す。最後に、 L_1 正則化付きロジスティック回帰を用いて二値識別器を学習し、特徴に対する重みベクトルを分析することにより数的優位な状況に至る過程および要因について考察を行う。

4.2 実験設定

実験では、データスタジアム株式会社 *1 が収集した 2015 年度の J リーグにおける試合のデータを使用した。選手のプレイログは手作業で、選手のトラッキングデータは各スタジアムに設置されたビデオカメラの映像から自動的に抽出され、一部手作業により修正が加えられたものである。トラッキングデータのサンプリング周波数は 25Hz である。3 チーム（浦和レッズ、サンフレッチェ広島、ガンバ大阪）について、各チーム 34 試合分のデータを用いた。今回の実験では、ゴールに比較的近い（15m 以内）位置からのシュートシーンのみを対象とし、3 チーム合わせて 503 回の攻撃機会について分析を行った。ロングシュートの成否は選手個人の能力による部分が大きいと判断し、今回の分析では除外した。

3.3 節で述べた二値識別器を構成する際に用いた特徴量を以下に示す。

1. プレイ履歴の時間間隔

*1 <https://www.datastadium.co.jp/>

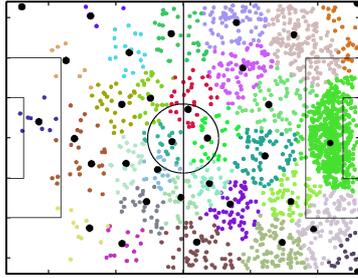


図 1: Mean Shift 法による代表点抽出の結果 .

表 3: パターン 4 における数的優位な状況の識別率 (%)

	L_1 -LR	ベースライン 1	ベースライン 2
識別率	85.9 (4.8)	72.1 (6.7)	79.2 (4.6)

2. ボールの移動距離
3. シュートした選手の走行距離
4. 攻撃側の走行距離 (一人当たり)
5. 守備側の走行距離 (一人当たり)
6. 攻撃側の全選手位置の中心 (x 座標)
7. 攻撃側の全選手位置の中心 (y 座標)
8. 守備側の全選手位置の中心 (x 座標)
9. 守備側の全選手位置の中心 (y 座標)

上記の 9 種類の特徴量をシュートからさかのぼって 3 ステップ前計算した . したがって , 特徴ベクトルの次元数は 27 である .

4.3 結果

4.3.1 攻撃パターンの抽出結果

図 1 に , ボールの位置情報に Mean Shift 法を適用することにより代表点を抽出した結果を示す . Mean Shift 法を実行する際に設定が必要なカーネル幅は 800 とした . 黒い点は代表点を表し , 色を付けた点は実際のボール位置を表す . 同じ色でプロットされたボール位置は , 同じ代表点に割り当てられたことを表す . 次に , 各攻撃機会に対して , ボールが各代表点を通った回数を計算することでボール軌跡をベクトルとして表現し , k -means 法を用いてクラスタリングを行った . ここで , $k = 5$ とした . 図 2 にクラスタリング結果を示す . サッカーのコート上に各攻撃パターンにおけるボール軌跡を描いた . 攻撃方向は右方向である . 灰色の円で表されているのは , ボール位置の代表点であり , 円が大きいほどボールが集まりやすいことを意味する . また , 矢印はボールの移動方向を表し , 矢印が太いほどパスやドリブルによって , ボールが移動しやすいことを意味する . 図 2 より , それぞれのクラスタが代表的な攻撃パターンを示していることが分かる . パターン 1 およびパターン 4 は両サイドからの攻撃 , パターン 3 およびパターン 5 はコーナーキックからの攻撃を抽出できている . パターン 2 はそれ以外の攻撃機会が含まれるクラスタであり , ボールの位置情報のみでは分離ができないような攻撃がノイズとして含まれていると考えられる .

4.3.2 数的優位を指標とする妥当性の検証結果

表 2 に , クラスタリングによって得られた各攻撃パターンに対するゴール率を示す . ゴール率は , シュートの回数に対するシュート成功の回数の割合である . 表 2 の二行目に示すように , 各攻撃パターンにおけるゴール率は , 0.05 から 0.26 の

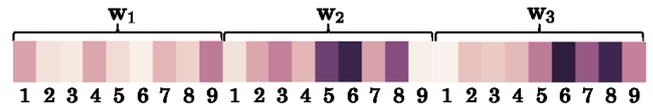


図 3: パターン 4 に対して L_1 正則化付きロジスティック回帰によって求められた重みベクトル .

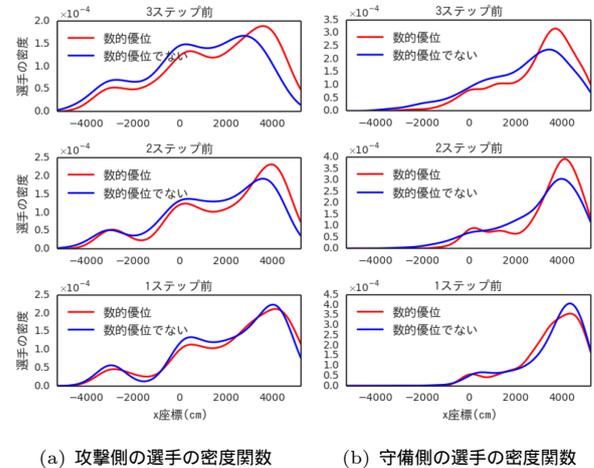


図 4: パターン 4 における各時間ステップの選手の密度関数 .

間であった . 次に , シュートの瞬間に数的優位な状況を作っていたか否かで攻撃機会を分類し , 二つの部分集合に対してゴール率を計算した . 表 2 より , パターン 1,3,4 において数的優位である場合とそうでない場合とで , シュートの成功確率に大きな差がみられた . パターン 2,5 については数的優位な状況を作ることが重要とは言えないことが考えられる . この結果から , 攻撃パターンに依って , 数的優位な状況を作ることが有効に働くシチュエーションとそうでない場合が存在することが分かる . パターン 5 のようにコーナーキックから開始される攻撃では , 守備側の選手が事前に守りを固めている場合が多く , 数的優位な状況を作りにくいことが考えられる . パターン 3 は , コーナーキックからの攻撃が多く含まれるが , コート中央からの攻撃も混ざってしまっているため , 数的優位な場合にシュート成功率が高くなったと考えられる . 一方で , パターン 1,4 のように , 選手がパスやドリブルをしつつ , センタリングによってボールをゴール前にいれるというような攻撃では , いかにも数的優位な状況を作り出すかが重要だと考えられる .

以上より , パターン 1,4 のような攻撃において , 数的優位な状況を作りだすことがシュート成功の確率を顕著に高めることが分かった . したがって , 数的優位な状況を作り出せたか否かという指標が , 攻撃の成否を評価するために妥当であると言える . 次節では , 数的優位か否かによってシュート成功率の差が最も大きかったパターン 4 に焦点を当てて分析を進める .

4.3.3 数的優位に至る過程の分析結果

パターン 4 に含まれるデータを用いて , 各攻撃が数的優位な状況を作り出した場合を正例 , そうでない場合を負例として , L_1 正則化付きロジスティック回帰 (L_1 -LR) を学習した . 学習の際の正則化パラメータは , 0.1, 0.5, 1, 10, 10^2 , 10^3 から 5 分割交差検定法によって決定した . 正例および負例のサンプル数は , それぞれ 20 および 83 である . L_1 正則化付きロジスティック回帰により学習された識別器の妥当性を評価するため , 二つのベースライン手法との識別率の比較を行った . ベースライン 1 は , 正例と負例の数に基づいて正例および負例のサン

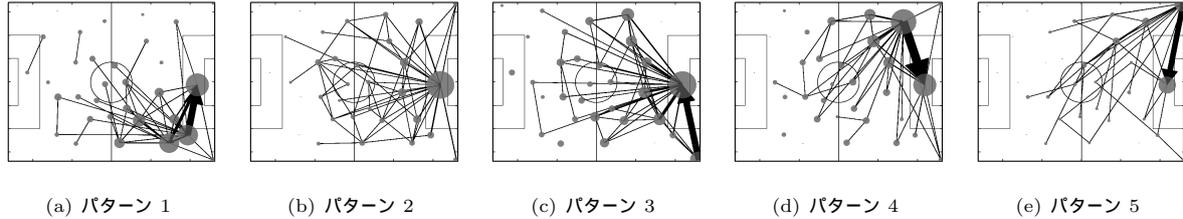


図 2: k -means 法によって抽出された各攻撃パターンにおけるボール軌跡の可視化結果 .

表 2: 各攻撃パターンにおけるゴール率

	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5
全体	0.187	0.250	0.256	0.165	0.052
数的優位	0.400	0.235	0.417	0.500	0.091
数的優位でない	0.152	0.255	0.229	0.084	0.043

ブル出現確率を計算し、その確率に基づいて識別を行うというものである。ベースライン 2 は、すべての攻撃機会を負例と識別する方法である。ベースライン 2 との比較を行う理由は、負例の数が正例の数に比べて非常に多いためである。表 3 に、5 分割交差検定法を用いて計算した各手法における識別率の平均値と標準偏差を示す。 t 検定 (p 値: 0.01) を行った結果、 L_1 -LR とベースライン手法との間に統計的優位差が認められた。この結果から、ベースライン手法より L_1 -LR の識別率が高く、 L_1 -LR の学習が妥当に行えていると言える。

次に、 L_1 -LR によって学習された重みベクトルを可視化した結果を図 3 に示す。重みの大きさを色の濃淡で表した。図 3 より、特に重みが大きかったのは w_2 および w_3 における特徴 6 (攻撃側の全選手位置の中心 (x 座標)) と特徴 8 (守備側の全選手位置の中心 (x 座標)) であった。この結果は、数的優位な状況を作り出すためには、シュートの直前よりも、その 2,3 ステップ前の行動 (パスやドリブル) の時点での、攻撃側と守備側の位置関係が重要であることを示唆している。図 4 にパターン 4 に含まれる各時間ステップの全選手の位置情報を用いて、選手の密度関数を描いた。上段から順に、シュートからさかのぼって 3, 2, 1 ステップ前の選手の密度関数を表す。密度関数はカーネル密度推定法 [Bishop 06] により計算した。密度関数を計算する際に必要なパラメータであるバンド幅は 0.25 とした。 x 軸は、コートの中心を 0 とし、左端を自チームのゴール位置 (-5250cm)、右端を相手チームのゴール位置 (5250cm) を表す。赤線はシュートの瞬間に数的優位な状況を作り出した全ての攻撃機会を用いて計算した密度関数であり、青線は数的優位な状況を作り出せなかった全ての攻撃を用いて計算したものである。攻撃側の密度関数が 3 つのピークを持つのは、選手のポジション (ディフェンダー、ミッドフィルダー、フォワードなど) によってプレイする位置がある程度決まっていることに起因する。各時間ステップにおける赤線 (数的優位) と青線 (数的優位でない) を比較する。1 ステップ前では、両者に大きな差はないが、2,3 ステップ前において、特にコートの前線 (x の値が大きいつき) において大きな差があることが分かる。特に、2,3 ステップ前における攻撃側の密度関数の前線におけるピーク位置が、青線よりも赤線の方が右にシフトしていることが分かる。このことから、シュートの時刻よりも 2,3 ステップ前の時点において、攻撃側の選手がいかにゴールに近い位置に移動できているか、ということが重要な要因であると考えられる。

5. おわりに

本研究では、サッカーにおける選手のプレイ履歴に基づいて、攻撃が成功に至る過程や要因についての分析を行った。実験では、ボールの軌跡から代表的な攻撃パターンを抽出可能とした。次に、攻撃パターンの種類に依るが、数的優位な状況を作り出すことがシュート成功の確率を顕著に高めることを示した。これにより、シュートの瞬間に数的優位な状況を作り出せたか否かが、攻撃の成否を評価する指標として妥当であることを確認した。最後に、各攻撃が数的優位な状況を作り出した場合を正例、そうでない場合を負例として、 L_1 正則化付きロジスティック回帰を学習し、学習された重みベクトルを分析することで、数的優位に至る過程や要因について考察を行った。今回の分析結果を踏まえて、今後はボール非保持者 (オフザボール) の動きに注目し、さらなる分析を行う予定である。

参考文献

- [Lucey 13] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews, "Assessing Team Strategy using Spatiotemporal Data," In *KDD'13*, pages 1366–1374, 2013.
- [Brooks 16] J. Brooks, M. Kerr, and J. Guttag, "Developing a Data-driven Player Ranking in Soccer using Predictive Model Weights," In *KDD'16*, pages 49–55, 2016.
- [Wei 13] X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan, "Large-scale Analysis of Formations in Soccer," In *DICTA'13*, 2013.
- [Wang 15] Q. Wang, H. Zhu, W. Hu, Z. Shen, and Y. Yao, "Discerning Tactical Patterns for Professional Soccer Teams: an Enhanced Topic Model with Applications," In *KDD'15*, pages 2197–2206, 2015.
- [Comanicu 02] D. Comanicu, and P. Meer, "Mean Shift: a Robust Approach toward Feature Space Analysis," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 24, no. 5, pages 603–619, 2002.
- [Tibshirani 96] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. of Royal Statist. Soc B*, vol. 58, no. 1, pages 267–288, 1996.
- [Bishop 06] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4. 2006: Springer New York.