

サッカーにおける守備側選手の動き評価手法で用いる
統計的判別モデルの高度化
Improvement of Statistical Classification Model
Used for Evaluating Defensive Player Movements in Soccer

大澤 賢人[†]
Kento Osawa

中田 洋平[†]
Yohei Nakata

1. はじめに

近年、様々なチームスポーツで、選手やボールの位置情報を用いた幾何学的フォーメーション解析やそこから得られる幾何学的フォーメーション特徴量を用いた研究が進展してきている[1]。しかし、著者らの調査の限り、守備側選手の動きの評価に応用した先行例は確認できなかった。そこで著者らは、幾何学的フォーメーション特徴量[1]から一定時間以内の失点確率を予測する統計的判別モデルを構築し、これに基づいて守備側選手の動きの適切さを評価する手法を提案してきた[2]。また、同手法を、実際の J1 リーグの選手・ボール位置情報に適用して、その有効性を定量的に評価し、考察してきた[3]。ただし、そこで用いられる統計的判別モデルは、あくまで数秒以内に被シュートが起こることを前提として失点確率を予測するものであり、被シュート自体の確率を予測するものではなかった。そこで、本稿では、評価対象を被攻撃シーン全体に拡張するためのデータ再整備を行い、被シュート確率も予測可能とする統計的判別モデルを新たに構築する。

2. 守備側選手の動き評価手法の概要

著者らがこれまで提案してきた守備側選手の動き評価手法[2][3]では、3種類の幾何学的フォーメーション特徴量[1]を用いる。1つ目は優勢領域ベース特徴量、2つ目は隣接グラフベース特徴量、3つ目は距離ベース特徴量である。優勢領域ベース特徴量は、文献[4]に示される運動モデルを考慮して算出される選手の支配領域に基づくものである。隣接グラフベース特徴量は、その支配領域の隣接関係に基づき、距離ベース特徴量は、選手と特定地点などとの距離情報に基づいて算出される。本手法では、このような幾何学的フォーメーション特徴量から対象時点 t から Δ 秒内の失点確率を予測する統計的判別モデルを学習する。次に、学習されたモデルを用いて、時点 t までのその動きの妥当性を評価する。具体的には、時点 t の δ 秒前から時点 t までに対象守備側選手が動けた範囲を、選手到達可能領域[4]に基づいて表現する。そして、時点 t において、同領域内の各代表点に同選手が移動したとした場合の失点確率を算出し、時点 t の実際の位置での失点の確率と比較することで、対象守備側選手の動きを評価する。

3. 統計的判別モデル高度化の概要

前述した守備側選手の動き評価手法[2][3]で用いてきた統計的判別モデルは、ロジスティック回帰モデル[5]であった。そして、このモデルでは、被シュートシーンに対して幾何学的フォーメーション特徴量[1]を入力とし、 Δ 秒以内に失点する確率を予測するように学習されていた。しかし、守

[†] 明治大学先端数理科学研究科

Graduate School of Advanced Mathematical Sciences,
Meiji University

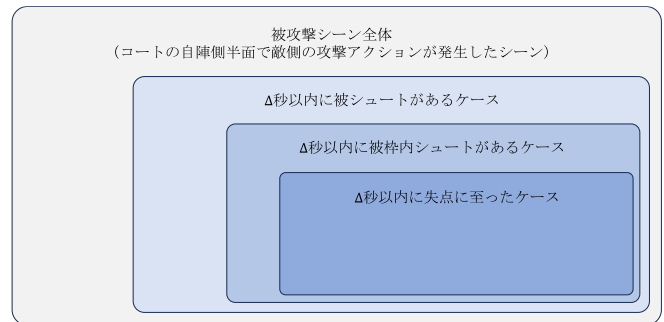


図 1 被攻撃シーンに関する包含関係

備には、失点を防ぐだけでなく、そもそも被シュートを防ぐ働きも含まれる。しかし、前述のモデルは、被シュート自体の確率を予測するものではなかったため、そうした守備の動きの適切さを評価することはできなかった。

そこで本稿では、まず評価対象を被攻撃シーン全体に拡大するため、被攻撃シーンに関する包含関係を整理した。図 1 にその概念図を示す。この包含関係に基づき、 Δ 秒以内に起きる事象として、(i) 被シュート無し、(ii) 被枠外シュート有り、(iii) 被枠内シュート有りだが失点無し、(iv) 失点有り、という 4 つの順序関係を定義した。そして、この順序関係を考慮したままで発生確率を予測可能なモデルとして、順序ロジスティック回帰モデル[5]を導入した。なお、本稿においては、被攻撃シーンとは、コートの自陣側半面で敵側の攻撃アクションが起きたシーンのことと定義した。ここで攻撃アクションとは、シュート、パス（スルーパス含む）、クロス、トラップ、ドリブル、フリックオンのことを指すものとした。ただし、後述するデータの関係上、ドリブルはドリブル開始時点での情報に基づくものとなる。

4. データの整備

前節に述べたような定義の下、データスタジアム社[6]から提供された 2016 年 J1 リーグの 45 試合のデータを用いて、順序ロジスティック回帰モデル[5]を学習するためのデータの整備を行った。具体的には、プレーログを参照して敵側の攻撃アクションの時点に対応する選手位置情報を抽出し、それらから著者らの先行研究[2][3]と同様の幾何学的フォーメーション特徴量を算出した。そして、それらを Δ 秒以内($\Delta=3$)に被シュートがあったケースとそれ以外のケースに分類した。なお、敵側のシュートがあった時点のものは、 Δ 秒以内に被シュートがあったケースとして扱う。更に、 Δ 秒以内に被シュートがあったケースについては、直後に起きた被シュートの結果に応じて前述の順序関係のラベル付けを行った。その上で、ラベル間のデータ数の不均衡を緩和するため、被シュート有り与被シュート無しの件数が等

表 1 各データ内での順序ラベル件数

順序ラベル	学習用	評価用
被シュート無	2080	245
被枠外有	836	85
被枠内有・失点無	1076	103
失点有	215	19

表 2 順位相関係数の算出結果

順位相関係数	学習用データ	評価用データ
スピアマン	0.700	0.671
ケンドール	0.543	0.526

しくなるよう、バランスドサンプリングを実施した。最後に、全 90 ハーフのうち、81 ハーフ分を学習用データとし、9 ハーフ分を評価用データとしてランダムに分割した。また、各特徴量は、学習用データ内における標本平均および不偏分散を用いて標準化を行った。表 1 には、各データ内における順序ラベルの件数を示す。

5. 順序ロジスティック回帰モデルの学習

前節で述べたように整備されたデータを用いて、順序ロジスティック回帰モデル[5]の学習を行なった。今回の学習では、入力とする特徴量の数が多いことから、ステップワイズ法(減増法)を用いて、特徴量選択を実施した。なお、その実行には R 言語を用いた。ただし、いくつかの特徴量については、学習が不安定となることから事前に排除した。そして、本研究で構築したモデルの予測性能を評価するために、被攻撃シーンの順序性をどの程度正確に捉えているかを検証した。具体的には、学習後のモデルから算出される予測スコア(線形予測子)と実際の順序ラベルとの関連性を評価した。評価には、スピアマンの順位相関係数及びケンドールの順位相関係数の 2 種類を用いた。

なお、スピアマンの順位相関係数は、両変数の値をそれぞれ順位に変換し、その順位間のピアソンの積率相関係数を計算することで、単調な関連の強さと方向を測る指標である。一方で、ケンドールの順位相関係数は、全てのデータのペアを取り上げ、両変数での順序関係に関して一致しているか、不一致であるかの数に基づいており、同様に単調な関連の度合いを示す。これらの相関係数はともに-1 から 1 の範囲の値を取り、1 に近いほど、モデルによる予測スコアと実際の順序ラベルとの間に強い順序的関連が存在することを示す。

表 2 は、学習用データと評価用データのそれぞれに対して、前述の 2 種類の順位相関係数を算出した結果をまとめたものである。学習用データに関しては、スピアマンとケンドールの順位相関係数のいずれにおいても中程度から強い正の相関が認められ、モデルが学習用データにおける被攻撃シーンの順序ラベルを比較的良好に捉えられていることが示唆される。また、評価用データに関しても同様に、いずれの順位相関係数も中程度からやや強い正の相関を示しているため、本モデルが未学習のデータに対しても適切に順序ラベルを予測できていることが確認された。

図 2 は、各順序ラベルの予測確率を曲線で示したもので

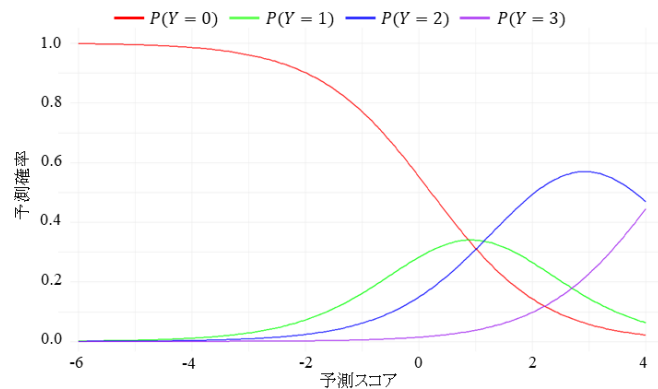


図 2 各順序ラベルの予測確率曲線

ある。 $P(Y=0)$ が被シュート無し、 $P(Y=1)$ が被枠外シュート有り、 $P(Y=2)$ が被枠内シュート有りだが失点無し、 $P(Y=3)$ が失点有りに対応する。横軸は予測スコアで、縦軸は各順序ラベルの予測確率を示しており、予測スコアの範囲は評価用データでのスコア値範囲に基づいている。同図からは、予測スコアが増加するにつれて、各順序ラベルの予測確率が期待通りに順序正しく滑らかに移り変わり、学習したモデルが適切に機能していることが視覚的に確認できる。ただし、同図の予測スコア範囲内では、失点有りの確率値 $P(Y=3)$ が他の順序ラベルの確率値より高くなる直前にスコア範囲が終了している。これは、本稿で用いたデータにおいては、失点が発生する場面自体が非常に少ないという性質が反映されたものと考えられる。

6. おわりに

本稿では、著者らが提案してきた守備側選手の動き評価手法[2][3]への導入を目的として、被攻撃シーン全体を対象としたデータ再整備を行い、被シュート確率も予測可能な順序ロジスティック回帰モデル[5]を学習した。その結果、モデルが適切に学習されていることを確認した。今後は、本稿で学習したモデルを用いて、評価手法[2][3]の高度化を図るとともに、特徴量の追加等を通じてモデルの更なる高精度化にも取り組む。

謝辞

データをご提供いただきましたデータスタジアム社[6]に深謝いたします。

参考文献

- [1] T. Mimura and Y. Nakada, "Enhancements of Pass Play Quantification Method with Geometric Features of Formations", Proc. ICEET 2022 pp. 1-6, (2022).
- [2] K. Osawa and Y. Nakada. "Defensive Movements Evaluation Method in Soccer Utilizing Geometric Formation Features and Statistical Classification Model", Proc. SCIS&ISIS 2024, pp. 1-2 (2024).
- [3] 大澤 賢人, 中田洋平, "サッカーにおける守備側選手の動きに対する評価手法の定量評価", 電子情報通信学会総合大会, A-10-03 (2025)
- [4] 藤村光, 杉原厚吉, "優勢領域に基づいたスポーツチームワークの定量評価", 電子情報通信学会論文誌, Vol.J87-D-II, No.3, pp. 818-828 (2004).
- [5] P. McCullagh, "Regression Models for Ordinal Data", Journal of the Royal Statistical Society: Series B (Methodological), Vol.42, Issue 2, pp.109-127 (1980).
- [6] データスタジアム株式会社: <https://www.datastadium.co.jp/>